DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

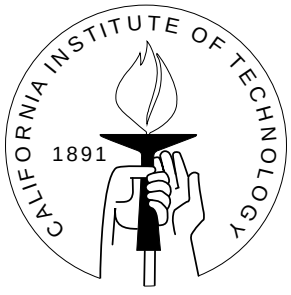
SOPHISTICATED EWA LEARNING AND STRATEGIC TEACHING IN REPEATED GAMES

Colin F. Camerer

Teck-Hua Ho
The Wharton School, University of Chicago

Juin-Kuan Chong
National University of Singapore

Forthcoming, *Journal of Economic Theory*



SOCIAL SCIENCE WORKING PAPER 1087R

April 2000
Revised July 2001

Sophisticated EWA Learning and Strategic Teaching in Repeated Games

Colin F. Camerer* Teck-Hua Ho Juin-Kuan Chong

Abstract

Most learning models assume players are adaptive (i.e. they respond only to their own previous experience and ignore others' payoff information) and behavior is not sensitive to the way in which players are matched. Empirical evidence suggests otherwise. In this paper, we extend our adaptive experience-weighted attraction (EWA) learning model to capture sophisticated learning and strategic teaching in repeated games.

The generalized model assumes there is a mixture of adaptive learners and sophisticated players. Like before, an adaptive learner adjusts his behavior the EWA way. A sophisticated player however does not learn and rationally best-responds to her forecasts of all other behaviors. A sophisticated player can be either myopic or foresighted. A foresighted player develops multiple-period rather than single-period forecasts of others' behaviors and chooses to 'teach' the other players by choosing a strategy scenario that gives her the highest discounted net present value. Consequently a foresighted player can develop a reputation for herself by strategic teaching if she is matched with an adaptive player repeatedly.

We estimate the model using data from p-beauty contests and repeated trust games. The generalized model is better than the adaptive EWA model in describing and predicting behavior. Including teaching also allows an empirical learning-based approach to reputation formation which is at least as plausible as the standard type-based approach.

1. INTRODUCTION

The process by which an equilibrium arises in a market or game has been a substantial mystery until recent years. Models of general equilibrium assume that equilibration comes from price-change rules implemented by a fictional Walrasian auctioneer (who is presumably a stand-in for some dynamic process which is typically unspecified). An implicit model of equilibration in game theory is that players figure out an equilibrium in a game, or adhere to a recommendation by an outside arbiter (perhaps a consortium of advisors or a government agency) if it is self-enforcing (e.g., Kohlberg and Mertens [39]). Biological models ascribe equilibration to genetic reproduction as well as mutation and natural selection. Early on, Nash spoke of a "mass action" interpretation of equilibration akin to natural selection (which is similar to modern accounts of cultural evolution).

None of these perspectives is likely to completely explain the actual time scale of equilibration in complex games played by humans. Humans learn faster than biological models predict, but not as fast (instantly!) as introspection. Therefore, a variety of other learning dynamics have been studied. Most studies ask about theoretical convergence properties of dynamics, primarily to see which equilibria they converge to (if any). This paper is about the empirical fit of learning models to experimental data. Our goal is to explain as accurately as possible, for every choice in an experiment, how that choice arose from a player's previous behavior and experience. We also strive to explain these choices using a general model which can be applied to any normal-form game with minimal customization.

The model we use extends the "experience-weighted attraction" (EWA) model of (Camerer and Ho [8], [9], [10]).[1] The key property of EWA is that it hybridizes features of popular learning rules, particularly reinforcement and belief learning (of the weighted fictitious play type), which have been widely studied in game theory. Hybridizing these familiar rules is useful for two purposes, one empirical and one theoretical. The empirical purpose is to fit and predict data better. Studies have found that the hybrid EWA typically improves substantially (and significantly) on reinforcement and belief models, in 31 data sets spanning a dozen different types of games (see details below). We are not aware of any learning model that has performed as well in that many statistical comparisons.

* This research was supported by NSF grants SBR 9730364 and SBR 9730187. Many thanks to Vince Crawford, Drew Fudenberg, David Hsia, John Kagel, and Xin Wang for discussions and help. Helpful comments were also received from seminar participants at Berkeley, Caltech, Harvard, Hong Kong UST, and Wharton.

[1] The model has also been applied to signaling games (Anderson and Camerer [2]), extensive-form centipede games (Camerer, Ho and Wang [11]) and bilateral call markets (Camerer, Hsia and Ho [12]).

The theoretical point of EWA is that belief learning and reinforcement learning are not different species; they are actually close cousins. When beliefs are formed according to weighted fictitious play, and used to calculate expected payoffs, those expected payoffs are exactly equal to a weighted average of previous payoffs, including "foregone payoffs" of strategies which were not chosen. Reinforcement models are averages (or cumulations) of previously received payoffs, excluding foregone payoffs. The *only* important difference between belief and reinforcement models is therefore the extent to which they assume players include foregone payoffs in evaluating strategies. In the EWA model, this difference is parameterized by a weight δ.[2]

This paper overcomes two important limitations of all adaptive models (including EWA and the special cases of reinforcement and belief learning).

One limitation is that adaptive players do not anticipate how others are learning and do not use knowledge of other players' payoffs (if they have it) to outguess their opponents. We add "sophistication" to the EWA model using two parameters. We assume a fraction α of players are sophisticated. Sophisticated players think that a fraction $(1-\alpha')$ of players are adaptive and the remaining fraction α' of players are sophisticated like themselves. They use the adaptive EWA model to forecast what the adaptive players will do, and choose strategies with high expected payoffs given their forecast. This 'self-consciousness' assumption creates a small whirlpool of recursive thinking which implies that standard equilibrium concepts (Nash), and sensible generalizations like quantal response equilibrium (QRE; McKelvey and Palfrey [43], [44]), are special cases of sophisticated EWA.

The idea of sophistication has been used before, in models of "level-k" learning (Ho, Camerer, and Weigelt [32], Stahl [66]; cf. Stahl and Wilson, [69]) and anticipatory learning (Selten [63]), although our parameterization is different. It shows that equilibrium concepts combine "social calibration" (accurate guesses about the fraction of players who are sophisticated, $\alpha = \alpha'$) with full sophistication ($\alpha = 1$). But these two features can be separated in principle, and it proves to be empirically useful to do so. The model is applied to data from p-beauty contest games (Nagel [49], Ho, Camerer and Weigelt [32]) and improves the fit substantially over purely adaptive models.

The second limitation of adaptive models is that they do not explain why behavior may depend on how players are matched with each other. Sophisticated players who are matched with the same players repeatedly may have an incentive to "teach" adaptive players, by choosing strategies with poor short-run payoffs which will change what adaptive players do, in

[2]When foregone payoffs are not known for sure, then elements of a set of possible payoffs or previously observed payoffs, can be used for 'payoff learning' (Anderson and Camerer [2], Camerer, Ho and Wang [11]).

a way that benefits the sophisticated player in the long-run. This kind of "strategic teaching" has been shown to give rise to repeated-game equilibria and reputation formation behavior through the interaction between "long-run" and "short-run" or myopic players (e.g., Fudenberg and Levine [25], Watson [72]). It also respects the fact that the way in which players are rematched makes an empirical difference in their behavior (they cannot "teach" in a protocol with no-repeat random matching, for example, and they do not appear to do so in experiments; e.g., Andreoni and Miller [3], Clark and Sefton [17]). We allow for teaching by adding a parameter ϵ to the sophisticated EWA model which represents the weight on future payoffs (like a discount factor). If $\epsilon = 0$, a player is sophisticated but does not incorporate the effects of current actions on future payoffs, i.e. she does not teach. If $\epsilon = 1$, the player fully accounts for the likely effects of current actions on future payoffs (as in standard repeated-game models).

We estimate the teaching model on data from experiments on repeated trust games. The model fits reasonably well although the data are noisy, and there is noticeable cross-session variation. It also exhibits the main patterns predicted by sequential equilibrium based on updating of entrants' beliefs about an incumbent's "type". Sophisticated EWA with strategic teaching therefore provides a boundedly rational model of reputation formation without the complicated apparatus of Harsanyi "types".

The next section describes the adaptive EWA model, motivates its structure, and briefly reviews earlier evidence. Section 3 introduces sophistication and shows empirical estimates from p-beauty contest games. Section 4 develops the teaching model and shows the empirical estimates from repeated trust games. Section 5 concludes.

2. ADAPTIVE EWA LEARNING

2.1. The Model

We start with notation. In n-person normal-form games, players are indexed by i ($i = 1, \ldots, n$). The strategy space of player i, S_i consists of m_i discrete choices, that is, $S_i = \{s_i^1, s_i^2, \ldots, s_i^{m_i-1}, s_i^{m_i}\}$. $S = S_1 \times \ldots \times S_n$ is the Cartesian product of the individual strategy spaces and is the strategy space of the game. $s_i \in S_i$ denotes a strategy of player i, and is therefore an element of S_i. $s = (s_1, \ldots, s_n) \in S$ is a strategy combination, and it consists of n strategies, one for each player. $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$ is a strategy combination of all players except i. S_{-i} has a cardinality of $m_{-i} = \Pi_{j=1, j\neq i}^{n} m_j$. The scalar-valued payoff function of player i is $\pi_i(s_i, s_{-i})$. Denote the actual strategy chosen by player i in period t by $s_i(t)$, and the strategy (vector) chosen by all other players by $s_{-i}(t)$. Denote player i's payoff in a period t by $\pi_i(s_i(t), s_{-i}(t))$.

EWA assumes each strategy has a numerical attraction, which determines the probability of choosing that strategy (in a precise way made clear below). Learning models require a specification of initial attractions, how attractions are updated by experience, and how choice probabilities depend on attractions. The core of the EWA model is two variables which are updated after each round. The first variable is $N(t)$, which we interpret as the number of 'observation-equivalents' of past experience relative to one period of current experience. (A player with a low $N(t)$ puts little weight on past attractions; a player with a huge $N(t)$ is barely affected by immediate experience.) The second variable is $A_i^j(a,t)$, an adaptive player i's attraction of strategy j <u>after</u> period t has taken place.[3]

The variables $N(t)$ and $A_i^j(a,t)$ begin with prior values, $N(0)$ and $A_i^j(a,0)$. These prior values can be thought of as reflecting pregame thinking or experience, either due to learning transferred from different games or due to introspection. (Then $N(0)$ can be interpreted as the number of periods of actual experience which is equivalent in attraction impact to the pregame thinking.)

Updating is governed by two rules. The first rule updates the level of attraction. A key component of the updating is the payoff that a strategy either yielded, or would have yielded, in a period. The model weights hypothetical payoffs that unchosen strategies would have earned by a parameter δ, and weights payoffs actually received, from chosen strategy $s_i(t)$, by an additional $1-\delta$ (so they receive a total weight of 1). Using an indicator function $I(x,y)$ which equals 1 if $x=y$ and 0 if $x \neq y$, the weighted payoff can be written as $[\delta + (1-\delta)\cdot I(s_i^j, s_i(t))]\cdot \pi_i(s_i^j, s_{-i}(t))$.

The parameter δ measures the relative weight given to foregone payoffs, compared to actual payoffs, in updating attractions. It can be interpreted as a kind of 'imagination' of foregone payoffs, or 'simulation' of outcomes under alternative competitive scenarios, or responsiveness to foregone payoffs. When δ is higher, players move more strongly, in a statistical sense, toward "ex post best responses".

The rule for updating attraction sets $A_i^j(a,t)$ to be the sum of a depreciated, experience-weighted previous attraction $A_i^j(a,t-1)$ plus the (weighted) payoff from period t, normalized by the updated experience weight:

$$A_i^j(a,t) \;=\; \frac{\phi\cdot N(t-1)\cdot A_i^j(a,t-1)}{N(t)} + \frac{[\delta + (1-\delta)\cdot I(s_i^j, s_i(t))]\cdot \pi_i(s_i^j, s_{-i}(t))}{N(t)}. \qquad (1)$$

[3]To prepare our notation for subsequent inclusion of sophistication, we use a in $A_i^j(a,t)$ to identify the attraction of an *adaptive* player; s is associated with a *sophisticated* player.

Note well that while we assume players are reinforced by their monetary payoffs, the reinforcement function could easily be altered to account for loss-aversion (the aversion to losses compared to equal-sized gains; cf. Tversky and Kahneman [70]), or for payoffs which reflect social preferences like fairness, reciprocity, inequality-aversion, and so forth (as in Cooper and Stockman [18]). One could also subtract an aspiration level or reference point (which may change) from payoffs, which is useful for keeping reinforcement models from getting "stuck" at nonequilibrium "satisficing" responses; EWA does something like this automatically, with no extra parameters.[4]

The decay rate ϕ reflects a combination of forgetting, and the degree to which players realize other players are adapting, so that old observations on what others did become less and less useful. When ϕ is lower, players discard old observations more quickly and are responsive to the most recent observations.

The second rule updates the amount of experience:

$$N(t) = (1-\kappa)\cdot\phi\cdot N(t-1)+1, \quad t \geq 1. \tag{2}$$

The parameter κ determines the growth rate of attractions, which reflects how quickly players lock in to a strategy.[5] When $\kappa = 0$, attractions are weighted averages of lagged attractions and past payoffs (with weights $\frac{\phi\cdot N(t-1)}{\phi\cdot N(t-1)+1}$ and $\frac{1}{\phi\cdot N(t-1)+1}$); so that attractions cannot grow outside the bounds of the payoffs in the game. When $\kappa = 1$ attractions cumulate, so attractions can be much larger than stage-game payoffs.

While we have not explicitly subscripted the key parameters δ, κ, and ϕ, they can obviously be different for different players or games (see Camerer, Ho and Wang [11]).

Attractions must determine probabilities of choosing strategies in some way. That is, $P_i^j(a,t)$ should be monotonically increasing in $A_i^j(a,t)$ and decreasing in $A_i^k(a,t)$ (where $k \neq j$). Three forms have been used in previous research: Exponential (logit), power, and normal (probit). We use the logit because it has compared favorably to the others in direct tests (Camerer and Ho [8]) and gracefully accomodates negative payoffs. The

[4] A strategy only increases in probability (holding previous attractions constant) if its payoff is above an average of the $\delta-$weighted foregone payoffs. Thus, EWA mimics a process in which reinforcements are payoffs minus an aspiration level which adjusts endogenously (reflecting foregone payoff).

[5] In our earlier papers (Camerer and Ho [8], [9], [10]), we define $\rho = (1-\kappa)\cdot\phi$ and call it the rate of decay for experience. The κ notation makes it clearer that the key difference is the extent to which attractions either average or cumulate.

logit form is:

$$P_i^j(a,t+1) = \frac{e^{\lambda \cdot A_i^j(a,t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(a,t)}}. \tag{3}$$

The parameter λ measures sensitivity of players to attractions. Sensitivity could vary due to the psychophysics of perception, whether subjects are highly motivated or not, or could reflect an unobserved component of payoffs (including variety-seeking, errors in computation, and so forth).

2.2. The EWA Learning Cube

Figure 1 shows a cube with axes representing the imagination parameter δ, the change parameter ϕ, and the lock-in parameter κ. Many existing theories are restricted cases of EWA learning which are represented by corners or edges of the cube. For example, cumulative reinforcement, average reinforcement, weighted fictitious play are edges and Cournot and fictitious play are vertices of this cube, as shown in the figure.

When $\delta = 0, \kappa = 1$ (and $N(0) = 1$), then $N(t) = 1$ and the attraction updating equation becomes $A_i^j(a,t) = \phi \cdot A_i^j(a,t-1) + I(s_i^j, s_i(t)) \cdot \pi_i(s_i^j, s_{-i}(t))$. This is the simplest form of cumulative choice reinforcement (Roth and Erev [58] and Erev and Roth [24]). When $\delta = 0, \kappa = 0$ (and $N(0) = 1/(1-\phi)$), the attraction updating equation becomes $A_i^j(a,t) = \phi \cdot A_i^j(a,t-1) + (1-\phi) \cdot I(s_i^j, s_i(t)) \cdot \pi_i(s_i^j, s_{-i}(t))$. This is a form of averaged choice reinforcement (attractions are averages of previous attractions and incremental reinforcement) (e.g., Mookerjee and Sopher [47]; cf. Sarin and Vahid [60]). The key property of reinforcement models is that they assume people ignore foregone payoffs. This simplifying assumption is most defensible in low-information environments in which players do not know much about the payoff landscape (although even then, more sophisticated rules might be used as players learn about foregone payoffs; e.g., Camerer, Ho, and Wang [11], Anderson and Camerer [2]). However, in most experimental games that have been studied empirically, players *do* know foregone payoffs and seem to respond to them. There is even evidence that pigeons are sensitive to foregone payoffs![6]

[6]Gallistel [28] (chapter 11) explains that the tendency of pigeons to "probability match" in binary choice experiments is affected by information about foregone payoffs. Specifically, pigeons tended to maximize, choosing one of two levers with the highest chance of delivering a reward all the time, when the pigeons knew after an unsuccessful trial that the other lever would have delivered a reward. (How did the pigeons "know"? Because a light displayed above a lever came on afterwards only if the lever had been armed for reward. If the light came on above the lever they did not choose, they 'knew' the foregone payoff.) When the pigeons did not know about the foregone payoffs (no light told them which lever had been armed to deliver food), they tended to "probability

A more surprising restricted case is weighted fictitious play (Brown [5], Fudenberg and Levine [26]).[7] When $\delta = 1, \kappa = 0$, then the attractions are updated according to $A_i^j(a,t) = \frac{\phi \cdot N(t-1) \cdot A_i^j(a,t-1) + \pi_i(s_i^j, s_{-i}(t))}{\phi \cdot N(t-1)+1}$. That is, attractions are weighted averages of lagged attractions and either realized or foregone payoffs. This sort of belief learning is a special kind of generalized reinforcement because beliefs can be written in the form of a difference equation. When beliefs are used to calculate expected payoffs for strategies, then the expected payoffs can also be written in the form of a difference equation: Expected payoffs are equal to previous expected payoffs and the increment in expected payoff which results from the updated belief. In the expected payoff equation, the belief disappears. The trick is that since beliefs are only used to compute possible future payoffs, and beliefs are backward-looking, possible future payoffs can be computed directly by incrementing expected payoffs to account for the "recently possible" foregone payoff. Seen this way, the difference between simple reinforcement and belief learning is a matter of degree, rather than kind (particularly the value of δ).

The parametric relation between belief and reinforcement models is subtle and went unnoticed for decades.[8] Why? One reason was that reinforcement theorists liked the idea of reinforcement precisely because it seemed to avoid "mentalist" constructs like beliefs. Furthermore, weighted fictitious play was first introduced in game theory as a heuristic way for players to reason their way to an equilibrium, not as a literal theory of how players lean from observation. It emerged from a way of thinking about learning that was (apparently) quite different from reinforcement.

Indeed, there is no compelling *empirical* reason to think parameter configurations which characterize human behavior will necessarily lie on the edges corresponding to belief and reinforcement learning, rather than on other edges or some interior regions. The kind of 'empirical privilege' that would justify focusing attention in those regions could have come from a variety of studies which continually show that measured parameters cluster in one portion of the cube. But that hasn't happened because nobody imagined the cube until recently. Most studies compare models from one corner or vertex with a static benchmark (usually Nash equilibrium). These

match" (to choose each lever about as often as that lever delivered a reward). So even pigeons notice foregone payoffs.

[7] Weighted fictitious play is a discrete dependent variable form of the adaptive expectations equation introduced by Cagan and Friedman in macroeconomics.

[8] For example, Selten [64] wrote "... in rote (reinforcement) learning success and failure directly influence the choice probabilities. ... Belief Learning is very different. Here experiences strengthen or weaken beliefs. Belief learning has only an indirect influence on behavior." EWA makes clear, however, that the indirect influence of learning of beliefs (for weighted fictitious play) can be exactly mimicked by direct influence.

studies provide little information about which learning rules– i.e., points in the cube– most accurately characterize how people learn.

2.3. Empirical Evidence

What parameters best characterize human learning? In previous empirical research, EWA has been used to fit and predict data from order-statistic coordination games (Camerer and Ho [9], [10]), p-beauty contests (Camerer and Ho [10]), mixed strategy games (Camerer and Ho [10]), n-person (Hsia [36]) and bilateral (Camerer, Hsia and Ho [12]) call-markets, cost allocation processes (Chen and Khoroshilov [16]), extensive-form centipede games (Camerer, Ho and Wang [11]),"unprofitable" games (Morgan and Sefton [48]), signaling games (Anderson and Camerer [2]), patent race games with iteratively dominated strategies (Rapoport and Amaldoss [55]), patent race games (Amaldoss [1]), and 5x5 matrix games (Stahl [68]).[9]

Table 1a-b summarize EWA parameter estimates and goodness-of-fit statistics from these 31 data sets. The goodness-of-fit statistic is -1 times log likelihood except in Chen and Khoroshilov [16]. The column "EWA" reports the $-LL$ of the EWA model. The reinforcement and belief models report the *difference* between the -LL's of those models and the EWA statistic. (Positive differences mean that EWA fits better.)

Values of δ tend to be between .5 and 1 in most studies except those in which games have only mixed-strategy equilibria, where δ is close to zero. The value of ϕ is reliably around .9 or so, with a couple of exceptions.

What about model comparisons? The fairest comparisons estimate parameters on part of a sample of data and forecast choices out-of-sample, so that models with more parameters will not necessarily fit better. (Indeed, if they succeed in-sample by overfitting, they will predict poorly out-of-sample.) In those 11 out-of-sample comparisons (denoted "OUT" in the third column from the right), EWA always outperforms reinforcement, although usually modestly. EWA outperforms belief learning in 9 of 11 cases, quite dramatically in some data sets.

Of course, EWA necessarily fits better in the other 20 in-sample comparisons than reinforcement and belief models which are special cases. One can use standard statistical techniques for penalizing more complex models – the χ^2 test, and Akaike and Bayesian criteria. These techniques are created so that if the special case restriction is true, the penalized fit of the more complex model will be worse than the fit of the restricted model. EWA generally does better even after applying these penalties. For example, if the difference in LL is 4 points or more then the special-case restriction will be rejected by the χ^2 test. By this criterion, EWA is more accurate

[9]Ho and Chong [33] applied the EWA model to fit and predict 130,000 consumer product purchases in supermarkets and found that EWA model fit substantially better than existing reinforcement models.

than belief learning in all in-sample comparisons, and more accurate than reinforcement in 16 out of 20 comparisons.

Figure 1 shows the locations of estimated parameter combinations from 20 games in Table 1[10] in the EWA cube. Each point represents a single game. The first observation is that points do not particularly cluster on the edges or corners corresponding to extreme-case theories, except for a few points in the lower corner corresponding to averaged reinforcement ($\delta = \kappa = 0$, and ϕ close to one). The second observation is that points are dispersed throughout the cube. Either learning rules are fundamentally different in different games– which creates the need for some theory of which parameter combinations are used in which games– or there may be some way to add something to the model to create more stability in parameter estimates.

Interestingly, three of four vertices on the $\kappa = 0$ and $\kappa = 1$ faces of the cube have been studied previously, but one has not. The fourth vertex, in which players are fully responsive to foregone payoffs ($\delta = 1$) but attractions cumulate rather than average past payoffs ($\kappa = 0$), does not correspond to any familiar learning theory (cf. McAllister [41]). However, the estimates from the three order-statistic coordination games are close to this segment. This vertex is also prominent in our estimates of the p-beauty contest game reported below. These results show an advantage of thinking about points in the learning cube. Parameter configurations which have never been combined previously turned out to characterize learning in some data sets better than models like belief and reinforcement learning which have been studied for fifty years.

We end this introductory section with a few comments. First, others have explored the econometric properties of EWA and some special cases of it. The news is not all good. For example, Salmon [62] finds with simulations that in 2x2 games, reinforcement, belief, and EWA models are often poorly recoverable in the sense that rule Y cannot be rejected as a good fit of data actually generated by a different learning rule X. EWA does least poorly in this sense because it does properly identify the value of δ. That is, when the data are generated by reinforcement (belief) models with δ=0 (=1), EWA model estimates are close to the correct value of δ. Blume et al. [4] find fairly good econometric performance of EWA and some other rules, when there are repeated samples and a substantial span of data. We regard these studies as harsh reminders that learning researchers should be more careful about investigating econometric properties of estimators before rushing to apply them. More constructively, pretests like this will

[10] In some studies, the same game was played at different stakes levels. In these cases, estimates were averaged across stakes levels.

help guide the choice of games and experiment lengths which are most likely to produce good econometric recovery.

Second, it would be useful to find interesting ways to economize on EWA parameters, creating a "EWA Lite". Since ϕ captures something about subjects' awareness of how their opponents are changing, δ captures subjects' sensitivity to foregone payoffs, and κ captures their desirability of locking in to a good strategy, making these parameters functions of observed history is a plausible way to go. Indeed, we have recently developed such an EWA Lite model and showed that it performs almost as well as the adaptive EWA model (Ho, Camerer, and Chong [31]). Excluding the initial conditions (which can be fixed by burning in the first-period data), the EWA Lite has only the payoff sensitivity parameter λ to be estimated. This one-parameter EWA Lite model should appeal to modelers who want a highly parsimonious model for describing learning behaviors in games.

Finally, it would certainly be useful to prove something about the long-run behavior of EWA players (cf. Hopkins [35]). Heller and Sarin [30] make a much-needed start in this direction. We conjecture that if $\kappa = 0$ (so that attractions are weighted averages of previous attractions and payoffs), then EWA players will converge to something akin to ϵ-equilibrium (at least in those classes of games where fictitious play converges) and ϵ will depend on δ and the payoffs in the game. The idea is that players could converge to a non-best response, but only if their stable payoff π_{stable} is greater than δ times the highest (best response) foregone payoff π_{br}. The gap $\pi_{br} - \pi_{stable}$ is a measure of ϵ. This may help explain why convergence in games with mixed-strategy equilibria is often so noisy, and such games are often a very poor way to distinguish models (e.g. Salmon [62]). Since mixed-strategy equilibria have knife-edge equilibria (the equilibrium mixtures are only weak best responses, by definition), the set of ϵ-equilibrium will often be much larger than the mixed equilibrium. Perhaps what we often observe in these games is players wandering among a large set of ϵ-equilibria which are produced by EWA equilibration.

3. EWA LEARNING WITH SOPHISTICATION

For game theorists steeped in a tradition of assuming players reason thoughtfully about the behavior of others, introducing sophistication into learning is a natural step; indeed, *not* assuming sophistication might seem strange. However, our standards are entirely empirical. We would like to know whether adding sophistication to an adaptive model parsimoniously (and the reverse, "dumbing down" sophisticated models by adding *un*sophistication) helps explain how people behave.

There are several empirical reasons to allow the possibility of sophistication:

1. Players *do* use information about others' payoffs. Several experiments that compare behavior with and without other-payoff information found a significant difference (Partow and Schotter [54], Mookerjee and Sopher [46], Cachon and Camerer [6]). The use of other-payoff information can also be tested directly, by measuring whether players open boxes on a computer screen that contain payoffs of other players. They do (Costa-Gomes, Crawford, and Broseta [19]; cf. Camerer, Johnson, Sen and Rymon [13]).

2. If players are sophisticated, the way in which they are matched when a game is played repeatedly can affect behavior. For example, compared to the random-opponent matching protocol, the fixed-opponent matching protocol should encourage players to adopt repeated game strategies.

3. Ho et al. [32] show that experienced subjects who played a second p-beauty contest converge significantly faster to Nash equilibrium than inexperienced subjects. This can be interpreted as evidence that players learned from the first p-beauty contest about how others were learning, which means they became increasingly sophisticated.

4. There is direct evidence in some games that players change strategies in ways which are inconsistent with adaptation, and consistent with sophistication. For example, Rapoport, Lo and Zwick [56] studied market entry games in which players had to enter one of three markets (see Ochs [52] for an overview). If a particular market was "under-entered", relative to the Nash equilibrium entry rate, then any sensible adaptive model (such as EWA and the restricted cases) predict more entry into that market in the next trial. In fact, players tended to enter even *less* frequently on subsequent trials, which is consistent with sophistication (i.e., expecting too much entry, and hence avoiding that market) rather than adaptation.

3.1. The Model

The population is assumed to consist of both adaptive learners and sophisticated players. We denote the proportion of sophisticated players by α and the proportion of adaptive players by $(1-\alpha)$. Adaptive learners follow the EWA learning rules and sophisticated players develop forecasts of others by assuming $(1-\alpha')$ proportion of the players are adaptive EWA learners and the rest are like themselves and best-respond to those forecasts.

Adaptive EWA learners follow the updating and probability equations (1)-(3). The sophisticated players have attractions and choice probabilities

specified as follows[11]

$$A_i^j(s,t) \;=\; \sum_{k=1}^{m_{-i}} [\alpha' P_{-i}^k(s,t+1)] \cdot \pi_i(s_i^j, s_{-i}^k) + (1-\alpha') \cdot P_{-i}^k(a,t+1) \quad (4)$$

$$P_i^j(s,t+1) \;=\; \frac{e^{\lambda \cdot A_i^j(s,t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(s,t)}}. \quad (5)$$

For a given player i, the likelihood function of observing a choice history of $\{s_i(1), s_i(2), \ldots, s_i(T-1), s_i(T)\}$ is given by:

$$\alpha \cdot [\Pi_{t=1}^T P_i^{s_i(t)}(s,t)] + (1-\alpha) \cdot [\Pi_{t=1}^T P_i^{s_i(t)}(a,t)] \quad (6)$$

The proposed model passes a form of the "publishability test" articulated by McKelvey and Riddough [45].They argue that a good social science theory should still apply even after it is "published" or widely-understood; or if behavior changes after publication, the theory should contain an explanation for why change occurs. Our model passes this test if only sophisticated players can "read", since sophisticated players will not change their behavior as long as adaptive learners remain unsophisticated. The only theory which passes full-readability test is when $\alpha = 1$ ($\alpha < 1$ corresponds to 'limited circulation' or 'illiterates'.)

Because the model assumes that sophisticated players think others are sophisticated (and those others think others are sophisticated...), it creates a whirlpool of recursive thinking which nests equilibrium concepts. Quantal response equilibrium (McKelvey and Palfrey [43], [44], Chen, Friedman and Thisse [15]; c.f. Rosenthal [57]) is equivalent to everyone being sophisticated ($\alpha = 1$) and the sophisticated players having rational expectations or "social calibration" about the proportion of sophisticates ($\alpha = \alpha'$). (Nash equilibrium, which we prefer to call 'hyper-responsive QRE', is just $\alpha = \alpha'$ along with infinite responsive sensitivity λ.) Weizsacker [74] allows players to have one response sensitivity, but think that other players' sensitivities are different (typically, lower) and finds that allowing this difference improves accuracy in explaining data from one-shot normal-form games.

Our parameterization emphasizes that QRE consists of the conjunction of two separate modeling assumptions: Players are sophisticated (α=1)

[11] This specification assumes that sophisticated players' and the modeler's forecasts of the adaptive players are identical. A more general specification can allow them to be different.

and sophisticated players are socially calibrated ($\alpha = \alpha'$). The two assumptions can be evaluated separately (and our model do so). Including both α and α' allows for two (opposite) kinds of judgment biases in assessing relative sophistication: Sophisticated subjects could underestimate the number of subjects who are sophisticated like themselves ($\alpha' < \alpha$, "false uniqueness" or overconfidence about relative sophistication), or could overestimate the number of sophisticates ($\alpha' > \alpha$, "false consensus" or "curse of knowledge").

Many studies document various types of optimism, overconfidence, or "false uniqueness" . For example, most people say they are above average on good traits and below average on bad traits. Most studies simply use self-reports and do not pay people according to their actual ranking, but a few studies have used experimental economics methods, which include financial incentive for accuracy and a clear definition of trait and rank, and replicate the basic finding. Applied to the sophisticated EWA model, overconfidence about relative sophistication would imply that sophisticates think there are fewer people as "smart" as themselves than there actually are, so $\alpha' < \alpha$. This kind of overconfidence is built into "level-k types" models like those of Stahl and Wilson [69]) (see also Costa-Gomes, Crawford, and Broseta [19], Ho et al. [32]). In those models, level 0 players choose randomly and level $k+1$ players best-respond to behavior of level k players. In a sense, this structure means players at every level think that nobody is as smart as them, and that everybody else is one level below. In our model, setting $\alpha' = 0$ corresponds to a level 1 learning type.

The opposite mistake is called "false consensus": People overestimate how much like themselves other people are.[12] A related effect is the inability of people who have learned new information to imagine what not knowing the information is like, the "curse of knowledge".[13] In sophisticated EWA a false consensus bias would imply that sophisticated people overestimate how many others are sophisticated, so that $\alpha' > \alpha$.

3.2. Dominance-solvable p-beauty contest games

We estimate the sophisticated WA model using data of the p-beauty contests collected by Ho et al. [32]). In a p-beauty contest game, n players

[12]The term "false consensus" is a misnomer because a 'rational bias' of this sort will result if people use their own behavior or tastes as a sample of tastes and update their beliefs about population tastes. The effect is correctly defined as overweighting your own tastes relative to information about the tastes of others. Engelmann and Strobel [23] show that there is no such bias when it is defined this way and information about other people is presented, along with financial incentives for accuracy.

[13]An example of false consensus is due to George Wu. He asked his students whether they had cell phones, and also asked them to estimate what fraction of students in the class had cell phones. Students with cell phones thought 55% had them, and those without cell phones thought only 35% had them.

simultaneously choose numbers x_i in some interval, say [0,100]. The average of their numbers $\bar{x} = \frac{\sum_i^n x_i}{n}$ is computed, which establishes a target number, τ, equal to $p \cdot \bar{x}$. The player whose number is closest to the target wins a fixed prize $n \cdot \pi$ (and ties are broken randomly[14]).

P-beauty contest games were first studied experimentally by Nagel [49] and extended by Duffy and Nagel [21] and Ho et al. [32]). These games are useful for estimating the number of steps of iterated dominance players use in reasoning through games. To illustrate, suppose $p = .7$. Since the target can never be above 70, any number choice above 70 is stochastically dominated by simply picking 70. Similarly, players who obey dominance, and believe others do too, will pick numbers below 49 so choices in the interval $(49, 100]$ violate the conjunction of dominance and one step of iterated dominance. The unique Nash equilibrium is 0.

There are two behavioral regularities in beauty contest games (see Nagel [50]). First, initial choices are widely dispersed and centered somewhere between the interval midpoint and the equilibrium. This basic result has been replicated with students on three continents and with several samples of sophisticated adults, including economics Ph.D.'s and a sample of CEOs and corporate presidents (see Camerer [7]). Second, when the game is repeated, numbers gradually converge toward the equilibrium.

Explaining beauty contest convergence is surprisingly difficult for adaptive learning models. Choice reinforcement converges far too slowly, because only one player wins each period and the losers get no reinforcement. Belief models with low values of ϕ, which update beliefs very quickly, may track the learning process reasonably well, but earlier work suggests Cournot dynamics are not fast enough either (Ho et al. [32]).

The sophisticated EWA model was estimated on a subsample of data collected by Ho et al. [32]. Subjects were 196 undergraduate students in computer science and engineering in Singapore. Each seven-person group of players played 10 times together twice, with different values of p in the two 10-period sequences. (One sequence used $p > 1$ and is not included below.) The prize was .5 Singapore dollars per player each time, about $2.33 per group for seven-person groups. They were publicly told the target number τ and privately told their own payoff (i.e., whether they were closest or not).

We analyze a subsample of their data with $p = .7$ and .9, from groups of size 7. This subsample combines groups in a 'high experience' condition (the game is the second one subjects play, following a game with a value of $p > 1$) and the 'low experience' condition (the game is the first they play).

[14]Formally, $\pi(x_i, x_{-i}) = \frac{n \cdot \pi \cdot I(x_i, argmin_{x_j} |x_j - \tau|)}{\sum_i I(x_i, argmin_{x_j} |x_j - \tau|)}$ where $I(x, y)$ is the indicator function that equals one if $x = y$ and 0 otherwise.

Several design choices were necessary to implement the model. The subjects chose integers in the interval [0,100], a total of 101 strategies. If we allow 101 possible values of $A^j(0)$ we quickly use too many degrees of freedom estimating the initial attractions. Rather than imposing too many structural requirements on the distribution of $A^j(0)$, we use the first-period data to initialize attractions.

Denote the empirically observed frequency of strategy j in the first period by f^j. Then initial attractions are recovered from the equations

$$\frac{e^{\lambda \cdot A^j(0)}}{\sum_k e^{\lambda \cdot A^k(0)}} = f^j, j = 1, \ldots, m. \tag{7}$$

(This is equivalent to choosing initial attractions to maximize the likelihood of the first-period data, separately from the rest of the data, for a value of λ derived from the overall likelihood-maximization.) Some algebra shows that the initial attractions can be solved for, as a function of λ, by

$$A^j(0) - \frac{1}{m}\sum_j A^j(0) = \frac{1}{\lambda} ln(\tilde{f}^j), j = 1, \ldots, m \tag{8}$$

where $\tilde{f}^j = \frac{f^j}{(\Pi_k f^k)^{\frac{1}{m}}}$ is a measure of relative frequency of strategy j. We fix the strategy j with the lowest frequency to have $A^j(0) = 0$ (which is necessary for identification) and solve for the other attractions as a function of λ and the frequencies $\tilde{f}^j$.

Since the subjects who did not win did not know the winning number, they could not precisely compute foregone payoffs. Therefore, we assume they have a simple uniform belief over a range of possible winning numbers.[15] We assume the losers reinforce numbers in the interval $[\tau - \frac{\delta \cdot n \cdot \pi}{d}, \tau + \frac{\delta \cdot n \cdot \pi}{d}]$. The amount of reinforcement is of a triangular form with the maximum of δ times the prize at the target number and decreases linearly at a slope of d (which is a parameter to be estimated). Denote the winning number to be w and the distance between the target and the winning number by $e = |\tau - w|$. Winners reinforce numbers in the intervals $(\tau - e, \tau + e)$ by δ times the prize. Winners reinforce the boundary number they choose, either $\tau - e$ or $\tau + e$, by the prize divided by the number of winners, and reinforce the other boundary number by δ times the prize divided by the number of winners. If there is only one winner, she also reinforce numbers in the intervals $(\tau - e, \tau - e - \frac{\delta \cdot n \cdot \pi}{d})$ and $(\tau + e, \tau + e + \frac{\delta \cdot n \cdot \pi}{d})$ by a reinforcement of a triangular form with the maximum of δ times the

[15] In Camerer and Ho [8], we assume that subjects know the winning number. Assuming subjects having a belief over the possible winning numbers provides a significantly better fit for the observed data.

prize at $\tau - e$ and $\tau + e$ and decreases linearly at a slope of d with smaller and larger number respectively.

Table 2 reports the results and the parameter estimates.[16] For inexperienced subjects, adaptive EWA generates Cournot-like estimates ($\hat{\phi} = \hat{\rho} = 0$ and $\hat{\delta} = .90$). Adding sophistication increases $\hat{\phi}$ and improves LL by 59.77 in-sample and 24.23 out-of-sample. The estimated fraction of sophisticated players is .24 and their estimated perception $\hat{\alpha}'$ is zero. The important consideration parameter δ is estimated to be .78 in the sophisticated model.

Experienced subjects show a large increase in sophistication. The estimated proportion of sophisticates, and their perceptions, are .77 and .41. Consequently, adaptive EWA fits much worse than sophisticated EWA, a difference in LL of 220.40 and 214.81 in- and out-of-sample. (In absolute LL terms, it also fits inexperienced and experienced subject data about equally well.) The increase in sophistication due to experience reflects a kind of "cross-period" learning which is similar to rule learning (Stahl [67]) or "rule switching" (Salmon [61]). The difference is that in Salmon and Stahl's approaches, players either keep track of actual or prospective performance of different rules, and switch in the direction of better-performing rules (Stahl [67]) or switch away from poorly-performing rules (Gale, Binmore and Samuelson [27], Salmon [61]). In our current specification, this change in rules can only occur between games, but it could be easily adapted to allow within-session rule changes (see Camerer and Ho [8]).

Figures 2a and 3a show the actual choice frequencies for inexperienced and experienced subjects, respectively. Experienced subjects actually start by choosing somewhat higher numbers (perhaps due to "negative transfer" from their earlier experience[17], but converge more rapidly.) By round 4 nearly half the experienced subjects choose numbers 1-10 (the large spikes on the back and left of Figure 3a). By contrast, it takes inexperienced subjects nine rounds for a third of them to choose 11-20. The fact that experienced subjects start *farther* from equilibrium, and end up much closer,

[16] We generate standard error for the parameter estimates using bootstrapping approach. 200 sets of bootstrapped estimates are produced using maximum likelihood estimation. Each set of bootstrapped estimates is derived from a weighted likelihood with randomly generated weights. One random weight is associated with each round of game. The random weights are non-negative integers generated such that their sum is equal to total rounds of games used in calibration (i.e. 8*number of sequences in in-sample calibration). For each parameter, we sort the 200 bootstrapped estimates and note the 2.5 and 97.5 percentile estimates (i.e. the 6th and the 295th estimates). We take the difference between these two estimates to be 3.92 standard errors.

[17] Experienced subjects had previously participated in a $p > 1$ beauty contest, in which choices converged to 200. This experience seemed to have a hysteresis effect on first-period choices.

is an indication that subjects are learning more rapidly, and more sophisticatedly (i.e., anticipating learning by others).

Figures 2b-c show the frequencies predicted by the adaptive EWA (2b) and sophisticated EWA (2c) models, for inexperienced subjects. Both models fit the general patterns in the data fairly well. Keep in mind that if the model overfits in periods 2-7, it would predict particularly badly in the out-of-sample periods 8-10, but it does not. The difference between Figures 2b and 2c shows that adding sophistication is only a small visual improvement (consistent with the modest increase in LL). The only noticeable improvement is that the largest spikes are higher in the sophisticated analysis, and closer to the data.

Figures 3b-c show the frequencies predicted by the adaptive EWA (3b) and sophisticated EWA (3c) models, for experienced subjects. Here there is a substantial improvement from including sophistication (compare 3c with 3b and 3a), which appears to fit nicely.

Restrictions of the sophisticated model generally degrade fit and predictive accuracy a lot. Imposing $\alpha' = \alpha = 1$ creates quantal-response equilibrium, which is 376.18 and 232.97 worse in LL for inexperienced and experienced subjects. One way for QRE to capture learning in a very reduced form is to allow the response sensitivity λ to vary over time. Allowing this *does* produce increasing values of λ (reported in Table 2), but the difference in in- sample LL is still very large, 365.19 and 165.77.[18] We suspect that the problem with QRE in these games is that the data are multi-modal, with "spikes" reflecting discrete levels of reasoning. (For example, in the very large samples from newspaper games with $p = 2/3$, there are sharp spikes at 33 and 22, reflecting one or two steps of iterated reasoning from a perceived mean of 50.) QRE will never produce spikes of this type and hence, will fit the central tendencies of the data, and track its movement toward Nash equilibrium over time, but will miss the multimodality. A QRE model in which different players have different values of λ can generate multimodality and will probably fit better, particularly in dominance-solvable games (in which increasing λ will correspond statistically to levels of iterated reasoning; cf. Goeree and Holt [29]).

[18] Our procedure of estimating the model in-sample and fixing parameter values to forecast out-of- sample makes life difficult for the varying-λ QRE model, since we fix λ at the last (period 7) estimate to forecast periods 8-10. A better procedure would impose some increasing functional structure on the $\lambda(t)$ function so that λ would continue to increase in the out-of-sample periods. Note, however, that the in-sample LL is substantially worse than sophisticated EWA for inexperienced subjects, and slightly worse with experienced subjects.

We did not estimate simple choice reinforcement models on these data because they do an awful job. Only one of seven players receives any reinforcement each period, so learning is far too slow.[19]

Cournot-like belief models will fit better; indeed, for both groups of subjects the EWA estimate of $\hat{\delta}$ is quite close to one and ϕ is low. However, as noted above, sophisticated EWA improves substantially on adaptive EWA, so even if adaptive EWA corresponds to belief learning, adding sophistication improves upon it.

We also estimated the rational expectations (RE, $\alpha = \alpha'$) and egocentric bias or level-type ($\alpha' = 0$) restrictions on estimated and perceived sophistication. For inexperienced subjects, the losses in log likelihood relative to sophisticated EWA are 3.63 and 2.85 for RE, in- and out- of-sample, and 21.24 and .16 for egocentric bias. For experienced subjects the analogous figures are 10.78 and 62.43 for RE, and 14.29 and 13.50 for egocentric bias.[20] While these differences in LL are modest, both restrictions can be rejected (particularly RE), which shows the predictive appeal of a model that separates sophistication and perceived sophistication, without imposing the strict level structure. In addition, the gap between α and α' grows with experience, from 24% to 34% (and the RE restriction is rejected much more strongly for experienced subjects). It seems that while players get more sophisticated between sessions, they also overestimate how many others become sophisticated.

A final observation provides an ideal segue to the next section of this paper. In the actual frequency plots Figures 2a and 3a, the curious eye can't help but notice the small number of very large choices (typically 100), particularly in later rounds. In Ho et al. [32], we called these "spoilers" and tested several explanations for why people might choose such high numbers. The most likely possibility is that subjects believe others are learning according to some adaptive rule that responds to the previous mean. By choosing a large number in round t, they throw adaptive learners off the trail, causing the adaptive learners to choose artificially high numbers in round $t+1$, which improves their chance of winning by choosing a low number. This kind of behavior combines two ingredients: A belief that others are learning adaptively; and a willingness to sacrifice period t profits (since picking 100 essentially guarantees a loss) for the sake of increased future profits, due to the way adaptive learners have been "taught" to behave. This is a first glimpse of strategic teaching.

[19] This is a general problem for choice reinforcement models in games where $n-1$ players earn nothing, such as auctions, winner-take-all tournaments, market games with one seller and many buyers (or vice versa), and so forth.

[20] Imposing these restrictions does not change other parameter estimates much, except to increase $\hat{\delta}$ to one for the egocentric restriction in both inexperienced and experienced subject samples.

4. STRATEGIC TEACHING

For a sophisticated player who anticipates that other players will learn, it is natural to take into account the effect of her period t action on the adaptive players' period $t+1$ actions, because those actions will change the sophisticated player's period $t+1$ payoffs. We call the behavior which maximizes discounted expected payoffs, taking into account the effect of one's own current behavior on future behavior of others, "strategic teaching".

The basic idea is described by Fudenberg and Levine [26] (pp. 261-263; cf. Ellison [22]). They write:

> ...imagine that one player is myopic and follows the type of learning procedure we have discussed in this book, while another player is sophisticated and has a reasonably good understanding that his opponent is using a learning procedure of this type. What happens in this case?... [much as] the results on equilibrium learning carry over to the case of nonequilibrium learning, so we expect that the lessons of the literature on reputation will carry over also to the case of nonequilibrium learning.

Fudenberg and Levine [25] showed that a very patient strategic teacher can get almost as much utility as from the Stackelberg equilibrium, by playing an optimal precommitment strategy forever (and waiting for the adaptive player to come around to best-responding). In their book they add (p. 262) that "the basic argument carries over in a straightforward way to the case of nonequilibrium learning" (cf. Watson [72], Watson and Battigali [73]).

Strategic teaching extends the reach of the EWA model to incorporate two important phenomena, which are beyond the grasp of standard adaptive models: (1) The influence of fixed-matching versus re-pairing protocols, and (2) emergence of repeated-game behavior including, importantly, reputation formation without cumbersome updating of "types" (a la Harsanyi).

If some players are capable of strategic teaching, then the way in which players are matched, and feedback they are given, should make a difference for learning. In fact, there is evidence that fixed-pair matching and random rematching produce different behaviors, which shows indirectly the likely influence of strategic teaching. For example, Andreoni and Miller [3] show that there is more mutual cooperation in finitely-repeated prisoners' dilemma games when subjects play repeatedly with a fixed "partner" than when they are re-paired with "strangers" in each period. Van Huyck, Battalio and Beil [71] found a similar phenomenon in two-player "weak-link games" (which are stag hunt or assurance games with seven strategies rather than two). They compared partner pairings with stranger re-pairing. The distributions of choices in the first period of the two pairing conditions were similar, but partner pairs were able to converge to the efficient equilibrium reliably (10 of 12 did so) while the stranger re-pairing behavior did not. Clark and Sefton [17] reported a similar result. It appears that

subjects who make efficient choices in the partner pairings, and see their partner choose an inefficient number in the first period, are inclined to "patiently" make an efficient choice once or twice more, as if holding their behavior steady and anticipating that the other player will learn to play efficiently.[21]

These stylized facts are consistent with strategic teaching. Strategic teachers who are matched with a different partner each time cannot use their current choices to influence what will happen in the future (to their benefit) if their future partners do not know the teachers' history of choices and anticipate similar choices in the future. Standard adaptive learning models do not directly predict differences across matching protocols.

Introducing teaching allows the possibility that repeated-game behavior is different than simply repeating stage-game behavior. Of course, in theory strategies which are not equilibrium strategies in a stage game can be used in repeated-game equilibria (by the threat of reversion to a Pareto-inefficient equilibrium if a defection occurs), as the "folk theorem" of repeated games would suggest. A familiar example is the repeated-PD, in which playing tit-for-tat is a repeated-game equilibrium (if the discount factor is large enough, relative to payoffs), supported by the threat of reversion to mutual defection, which is less profitable. This kind of dynamic is precisely what teaching can explain: A strategic teacher may play a strategy which is not myopically optimal (such as cooperating in a PD) in the hope that it induces adaptive players to expect that strategy in the future, which triggers a best-response that benefits the teacher. Furthermore, reversion to the Pareto-inefficient equilibrium is credible because the teacher knows that if she defects, her adaptive opponent will learn to quit playing the repeated-game strategy.

Strategic teaching is a different way to comprehend repeated-game equilibria than standard analyses, and might prove *better* as a way of explaining actual behavior. Consider the influence of the length of the horizon of future play. In standard (pre-1980) theory, folk theorem results unravel when the horizon of the repeated game is finite. Standard theory therefore cannot easily explain why players cooperate in say, the first 22 periods of a 25-period repeated PD, as is typically observed in experiments (e.g., Selten

[21] The same difference in partner and stranger matching does *not* seem to be present in three-player groups (see Knez and Camerer [38]). We conjecture that the difference in two− and three−player dynamics can be traced to strategic teaching. In both cases, the success of a strategic teacher who makes the efficient choice repeatedly depends on behavior of the "least teachable" player. Since there are two other players being taught in the three-player game, and only one in the two-player game, strategic teachers are more likely to give up and converge toward inefficiency in three-player games. The same sort of dynamics might help explain the fact that collusion is more sustainable in repeated pricing and quantity games when the number of players is small, and difficult to sustain when the number is large (e.g., Holt [34]).

and Stoecker [65]). Since strategic teaching assumes that the learners are adaptive, and do not use backward induction, strategic teaching will generally predict folk theorem-type results until a point near the end of the finite horizon, when it no longer pays to teach because the end is too close. Strategic teaching therefore does *not* predict unraveling in finitely-repeated games, which is consistent with most experimental data and everyday intuition (and contrary to standard theory).

Of course, it is now well-known that repeated-game behavior can arise in finite-horizon games when there are a small number of "irrational" types (who essentially act like the horizon is unlimited), which creates an incentive for rational players to behave as if the horizon is unlimited until near the end (e.g., Kreps and Wilson [40]). However, specifying why some types are irrational, and how many they are, makes this interpretation difficult to test.[22] In the teaching approach, the "crazy" type the teacher wants to pretend to be arises endogenously from the payoff structure– they are generally Stackelberg types, who play the strategy they would choose if they could commit to it. (Teaching substitutes for "true" commitment.) In trust games, they would like to commit to behaving nicely; in entry-deterrence, they would like to commit to fighting entry.

4.1. The Model

To illustrate the details of how teaching works, consider the repeated trust game. (Below, we estimate the teaching model on a sample of experimental data for this game, from Camerer and Weigelt [14].) In the trust game, there is an borrower B who want to borrow money from a number of lenders denoted L_i $(i = 1, \ldots, N)$. A lender makes only a single lending decision (either *Loan* or *No Loan*) and the borrower makes a string of N decisions each of which, either (*repay* or *default*), is made after observing the lender's decision.

In a typical experimental session, a group of subjects are randomly assigned the role of borrower, or lender. For example, a group of 11 subjects may be divided into a group of 3 borrowers and a group of 8 lenders. In a single sequence, an borrower B is randomly chosen to play in an 8-round supergame. Each lender $L_i (i = 1, \ldots, 8)$ plays in 1 of the 8 stage games in a random order (which is unknown to the borrower). To study cross-sequence learning, the entire supergame is repeated in a series of sequences.

Denote each sequence of game rounds by k and each game round by t. Note that within the sequence of game rounds, there is a common borrower. In a typical experimental session, there are about 81 sequences. The goal

[22] Some models allow the number and nature of irrational types to be a free parameter, as in the "homemade prior" account of Camerer and Weigelt [14] and Palfrey and Rosenthal [53], executed formally by McKelvey and Palfrey [42]. The agent-based QRE model we use below incorporates this idea.

is to specify the probabilities for both the borrower and the lender for each of their actions, in each round of each sequence.

Recall that each lender plays only once, and each borrower plays in only a third of the sequences. Yet they watch all the other plays, and clearly respond to observed behavior of others. Therefore, we assume a kind of "observational" learning; all lenders learn equally from what they observe. While the common-learning assumption is a substantial departure from other sorts of adaptive learning models, it is plausible and well-supported by some experimental evidence (e.g., Duffy and Feltovich [20]) and is necessary to explain what we see in these games.[23] We assume that the lenders are purely adaptive (because the matching scheme gives them no incentive to teach) and the borrower may be sophisticated, and may also be a strategic teacher.

An important twist in our model is that players are assumed to learn about the attraction of a strategy in a current round in two separate ways: They learn from previous rounds in a sequence; *and* from how the strategies performed in the current round in previous sequences. For concreteness, consider round 7 in sequence 14. A lender presumably sees what happened in the previous 6 rounds, and learns about whether to loan from what happened in those rounds. It is also plausible that the lender looks at what happened in the 7th round of previous sequences 1-13, and learns about whether she should loan in round 7 from those sequences.

We include both types of learning in the model for theoretical and empirical reasons. The theoretical reason is that learning about a specific round across sequences is like repeated-stage-game learning across similar games; where the "similar" games are identical rounds in previous sequences. This sort of transfer has been explored by Stahl [67] and resembles the similarity-based "spillover" of reinforcement from a chosen strategy to neighboring strategies explored by Sarin and Vahid [59]. Empirically, we thought that including cross-sequence learning was necessary to explain the data better. After all, the reason why experimenters conduct a (long!) series of repeated game sequences, rather than simply one, is presumably a prior belief that learning required many repetitions of the entire sequence. This sort of cross-sequence learning is precisely what our two-step learning process (to be discussed below) allows. The strength of cross-sequence learning is parameterized by a single parameter τ. If that parameter is zero there is no cross-sequence learning. So the data can tell us whether allowing cross-sequence learning is helpful through the value of τ.

It is not clear how to integrate the two sorts of learning. Returning to our example, the strategy Loan for a lender before period 7 of sequence

[23]That is, lenders in later rounds of a sequence clearly react to what happened in earlier rounds, even though the earlier-round behavior did not affect their payoffs. To explain this sort of behavior without observational learning is impossible.

14 can be said to have two different attractions - the attraction of Loan after period 6, and the attraction of Loan after period 7 of sequence 13. Simply averaging these attractions is an obvious, but hamfisted, way to include them both in a general learning process. Reflecting a prior belief that within-sequence learning is more important than cross-sequence learning, we elected to make updating attractions within a sequence the basic operation, then include an extra step of partial updating using the average payoff from previous sequences.

Let us specify the attraction of an adaptive lender at the end of sequence k and round t for strategy j, $A_L^j(a,k,t)$. The updating occurs in 2 steps. The idea is to create an "interim" attraction for round t, $B_L^j(a,k,t)$, based on the attraction $A_L^j(a,k,t-1)$ and payoff from the round t, then incorporate experience in round $t+1$ from previous sequences, transforming $B_L^j(a,k,t)$ into a final attraction $A_L^j(a,k,t)$.

- Step 1 (adaptive learning across rounds within a sequence):

$$\begin{aligned} B_L^j(a,k,t) &= \frac{\phi \cdot N(k,t-1) \cdot A_L^j(a,k,t-1)}{M(k,t)} + \\ &\qquad \frac{(\delta + (1-\delta) \cdot I(j, s_L(k,t))) \cdot \pi_L(j, s_B(k,t))}{M(k,t)} \\ M(k,t) &= \phi(1-\kappa) \cdot N(k,t-1) + 1 \end{aligned}$$

- Step 2 (simulated learning in a coming round from previous sequences):

$$\begin{aligned} A_L^j(a,k,t) &= \frac{\phi^\tau \cdot B_L^j(a,k,t) \cdot M(k,t) + \tau \cdot \delta \cdot \hat{\pi}_L^j(k,t+1)}{N(k,t)} \\ N(k,t) &= [\phi(1-\kappa)]^\tau \cdot M(k,t) + \tau \end{aligned}$$

We assume that the learning about an upcoming round from previous sequences is driven by the average payoff in that round in previous sequences. Formally, $\hat{\pi}_L^j(k,t+1) = \sum_{m=1}^{k-1} \pi_L(j, s_B(m,t+1))/(k-1)$.[24] As usual, we derive $P_L^j(a,k,t+1)$ from $A_L^j(a,k,t)$.

Next we specify learning by an **adaptive** borrower. The updating occurs in 2 steps.

- Step 1 (adaptive learning across rounds within a sequence):

$$B_B^j(a,k,t) = \frac{\phi \cdot N(k,t-1) \cdot A_B^j(a,k,t-1)}{M(k,t)} +$$

[24] We have also explored a specification in which only the payoff received in the previous sequence from a particular strategy is used. That is, $\hat{\pi}_L^j(k,t+1) = \pi_L(j, s_B(k-1,t+1))$. That specification is too "fickle" and fits worse than the average-payoff specification.

$$\frac{(\delta + (1-\delta)\cdot I(j, s_B(k,t)))\cdot \pi_B(j, s_L(k,t))}{M(k,t)}$$

$$M(k,t) = \phi(1-\kappa)\cdot N(k,t-1) + 1$$

- Step 2 (simulated learning in a coming round from previous sequences):

$$A_B^j(a,k,t) = \frac{\phi^\tau \cdot B_B^j(a,k,j)\cdot M(k,t) + \tau\cdot\delta\cdot\hat{\pi}_B^j(k,t+1)}{N(k,t)}$$

$$N(k,t) = [\phi(1-\kappa)]^\tau \cdot M(k,t) + \tau$$

As above, we assume that the learning from an upcoming round, from previous sequences, is driven by the average payoff in that round in previous sequences ($\hat{\pi}_B^j(k,t+1) = \sum_{m=1}^{k-1}\pi_B(j, s_L(m,t+1))/(k-1)$). We derive $P_B^j(a,k,t+1)$ from $A_B^j(a,k,t)$.

Now we are ready to specify how a **sophisticated** borrower will behave. A sophisticated borrower guesses how the lender learns, and adapts those guesses to experience, and also plans actions for the remaining periods within a game sequence. Specifically, we assume a sophisticated borrower's attractions are specified as follows:

$$A_B^j(s,k,t) = \sum_{j'=Loan}^{NoLoan} P_L^{j'}(a,k,t+1)\cdot\pi_B(j,j') +$$

$$\max_{J_{t+1}}\{\sum_{v=t+2}^{T}\epsilon^{v-t-1}\sum_{j'=Loan}^{NoLoan}\hat{P}_L^{j'}(a,k,v|j_{v-1}\in J_{t+1})\cdot\pi_B(j_v\in J_{t+1},j')\}$$

where $\hat{P}_L^{j'}(a,k,v|j_{v-1}) = \hat{P}_L^{Loan}(a,k,v-1|j_{v-1})\cdot P_L^{j'}(a,k,v|(Loan,j_{v-1})) + \hat{P}_L^{NoLoan}(a,k,v-1|j_{v-1})\cdot P_L^{j'}(a,k,v|(NoLoan,j_{v-1}))$. J_{t+1} specifies a possible path of future actions by the sophisticated borrower from round $t+1$ until end of the game sequence. That is $J_{t+1} = \{j_{t+1}, j_{t+2}, \ldots, j_{T-1}, j_T\}$ and $j_{t+1} = j$. We search only paths of future actions that always have default following repay because the reverse behavior (repay following default) generates less return; the latter kind of behavior has a smaller reputation building effect on the lender's behavior. It is therefore in the sophisticated borrower's interest to allocate repay to earlier rounds. As usual, we derive $P_B^j(s,k,t+1)$ from $A_B^j(s,k,t)$ using a logit rule.

Note that if $\epsilon = 0$, the player is sophisticated but myopic (she does not take into account the future learning effects of current actions). If $\epsilon > 0$,

the sophisticated player is a teacher who takes into account the effects of current actions on learned behavior of others.

As before, we can assume that a proportion α of the borrowers are sophisticated. Then the likelihood of observing the data is given by $\Pi_k[(1-\alpha)\cdot \Pi_t P_B^{S_B(t)}(a,k,t) + \alpha \cdot \Pi_t P_B^{S_B(t)}(s,k,t)]$.[25]

We estimate the model using repeated game trust data from Camerer and Weigelt [14]. As in our earlier work, we use maximum likelihood estimation (MLE) to calibrate the model on about 70% of the sequences in each experimental session, then forecast behavior in the remaining 30% of the sequences. If the model fits better in-sample by overfitting, it will perform surprisingly poorly out-of-sample.[26]

4.2. Repeated Trust Game

We fit the teaching model on experimental data on the repeated borrower-lender trust game studied by Camerer and Weigelt [14], to see whether the model can explain borrower teachers trying to reassure worried lenders by paying back in early periods, to benefit both sides.

Table 3 shows payoffs in the repeated trust game. The lenders earn 10 if they do not lend; they earn 40 if a loan is repaid and lose -100 if the borrower defaults.[27] A normal borrower earns 10 if the lender does not lend, 150 if the lender lends and she defaults, and earns only 60 if she pays back. Honest-type borrower have default and repayment payoffs of 0 and 60 respectively (note that they earn more from repaying).

The probability that a borrower had honest-type payoffs in a particular sequence, P(Honest), was .33 (sessions 3-5), .10 (sessions 6-8) and 0 (sessions 9-10). Subjects were MBA students at NYU or University of Pennsylvania. They were paid according to performance and earned an average of $18 for a 2-1/2 hour session.

Each session had 70-101 eight-period sequences. In each session, there were 11 subjects, three borrowers and eight lenders, in a fixed-role protocol. In a given eight-period sequence, a single borrower faced eight different lenders. To inhibit two-sided reputation-building, the eight lenders played in a different random order in each sequence. To prevent borrowers from

[25] We assume rational expectation (i.e., $\alpha = \alpha'$) in the estimation of the teaching model.

[26] We used GAUSS. To avoid settling into local maxima, we posited two or three starting values for each parameter, and used 64 combinations of possible parameter values as different initial conditions. After 50 iterations from each initial condition, we chose the best-fitting estimates and continued iterating to convergence.

[27] Payoffs were varied for lenders for the loan-default outcome, -50 in sessions 6-8 and -75 in sessions 9-10. These parameter variations provide a small 'stress test' for whether the same structural model can account for behavior across sessions with minimal parameter variation.

treating consecutive sequences as a long cross-sequence supergame, the lender chosen to play in sequence $k+1$ was one of the two subjects who did not participate in sequence k (i.e., lenders never played two sequences in a row).

We now discuss the sequential equilibrium predictions, then return to the data. With the payoffs used in Table 3, the analysis proceeds as follows: Start from period 8. In this period, lenders know that the borrower will play Default if she loans (and the honest type will, of course, repay) so the only question is the probability that the borrower is a honest type. Simple algebra shows that if lenders are risk-neutral, they should loan if P(Honest) is above 55/70, about .79. Define this important threshold to be γ. In the second-to-last period, period 7, normal borrowers are torn between two forces: Conditional on loan, they would like to choose Default to earn the higher payoff; but if they do so, they would have revealed their type and would earn the No-Loan payoff in period 8. However, if their reputation (i.e., the perception P(Honest) lenders have) in period 7 is below γ, then lenders will not loan and Bayesian updating would lead lenders to have the same perception in period 8 which, by assumption, is too low to induce the lenders to lend in period 8. The trick is for borrowers to play a mixed strategy, repaying frequently enough that if they do default, the updated P(Honest) will be just above γ, so that lenders would loan in the last period.

Given a particular P(Honest) in period 7, the normal borrower should choose a repayment probability p which keeps the lender indifferent to loan in period 7, and allows Bayesian updating of his reputation to the threshold P(Honest)=γ in period 8. Combining these two conditions gives a threshold of perceived P(Honest) which happens to be γ^2, and a mixed strategy probability of loan in period 7 of .560.

The same argument works by induction back to period 1. In each period the lender has a threshold of perceived P(Honest) which makes her indifferent to loan. The path of these P(Honest) values is simply γ^n. Figure 4 shows this path, and the mixed-strategy probabilities of paying back by normal borrowers which keep the lender's perceptions along this path (for an initial prior P(Honest) of .33). The Figure can be used to illustrate all the key properties of this equilibrium.[28] In the first three periods, the threshold P(Honest) is below the prior of .33, so borrowers can "afford" to always default and lenders should loan. Beginning in period 4, normal borrowers must mix in order to boost their reputation, conditional on loan, to stay along the equilibrium path of P(Honest) which increases. If

[28]Characteristically, there are other sequential equilibria. For example, the normal borrower might never repay, if she thinks that the lender will perceive Repay as an indication of a normal type. The intuitive criterion selects the equilibrium we discuss however, so we will casually refer to it as "the" equilibrium.

the borrower ever defaults, the lender should not lend in all subsequent periods.

Two patterns in the data are of primary interest. First, what is the rate of loan across periods (and how does it change across sequences)? Second, how do borrowers respond to loan in different periods (and how do these responses vary across sequences)?

Typical patterns in the data can be seen in Figures 5a-b. Figures 5a-b are data from all sessions. The figures show relative frequencies of No Loan and Default (conditional on loan).[29] Sequences are combined into ten-sequence blocks (denoted "sequence") and average frequencies are reported from those blocks. Periods 1,...,8 denote periods in each sequence.

Figure 5a-b show the data from all eight sessions pooled. Lenders start by generally making loans (i.e., low frequency of no-loan) in early periods, then learn to rarely loan in periods 7-8. Borrowers default infrequently in early periods, but usually default in periods 7-8, and the pattern is particularly dramatic in later sequences.

As a benchmark alternative to the teaching model, we estimated an agent-based version of quantal response equilibrium suitable for extensive-form games (see McKelvey and Palfrey [44]). An Appendix explains precisely how the the agent-QRE model is implemented and estimated.[30] We use this model, rather than sequential equilibrium, because the (intuitive) sequential equilibrium predicts many events to have zero probability, so some notion of error or trembling is needed to fit the data (otherwise the logarithm of likelihood explodes). Agent-QRE is a plausible form and fits many data sets well (see McKelvey and Palfrey [44], Goeree and Holt [29]). We implement the model with four parameters – the prior belief of lenders about P(Honest) (which can differ from the prior induced by the experimental design to reflect "homemade priors"), and different response sensitivities λ for lenders, honest borrowers, and normal borrowers. Agent-QRE is a good benchmark because it incorporates the key features of repeated-game equilibrium – (stochastic) optimization, accurate expectations about actions of other players, and Bayesian updating. Also, while it makes the same conditional predictions in every sequence, it can explain differences across sequences if early-period play changes across an experimental session. AQRE therefore presents a stiff challenge to any adaptive learning model which tries to explain learning both within and across sequences.

[29] To distinguish cases in which there are no data from true zeros (e.g., no repay after several loans), we plot cases with no data as -.1.

[30] We use an agent-based form in which players choose a distribution of strategies at each node, rather than using a distribution over all history-dependent strategies.

Table 5 gives parameter estimates for each of the 8 sessions, estimated separately to gauge cross-session stability. Table 4 shows measures of fit.[31] To measure fit we report both log likelihood and the average predicted probability of events that occurred. Table 4 shows that for the in-sample calibration, the average predicted probabilities range from 59% to 83% for the teaching model, compared to 54% to 82% for agent-QRE, and the teaching model fits better in every session. Of course, an important test for overfitting is how badly performance degrades in out-of-sample forecasting. The average probabilities fall by only 0.25% on average for the teaching model, and range from 59% to 86%. The agent-QRE model has average probabilities range from 54% to 85% and always fits worse than the teaching model by this measure. The log likelihood measure yields a similar result: The teaching model beats agent-QRE in all but one session.

Table 5 shows parameter values (and standard errors) for the teaching model. The interesting lender parameters are δ and τ. The weights on foregone payoff, δ range from .18 to .65, and average .43, with low standard errors (generally less than .10). Estimates of τ range from .88 to 1.00, except for one outlier at .51, which indicates a high degree of cross-sequence learning.

The interesting parameters for sophisticated borrowers are α and ϵ. The degree of sophistication α ranges from .03 to .99. The "horizon" parameter ϵ is close to one in five sessions and around .20 in three sessions.[32]

An important test of the agent-QRE model is whether the estimated perceived P(Honest) is reasonably correlated with the induced prior P(Honest) across sessions. The estimated priors in the eight sessions are .54, .19, .07, .42, .41, .50, .00 and .00 (these estimates are precise: standard errors are no larger than .004). The correlation of these estimates and the induced priors across the eight sessions is only .23. This suggests that the agent QRE model is *not* capturing a behavioral process in which the induced prior plus some stable behavioral disposition toward always repaying explains what players are doing.

Figures 5c-d show that the teaching model captures most of the key regularities in the data, although it does not always fit aggressive patterns in the data. No-loan frequencies are predicted to start low and rise across periods, as they do. There are some slight downward trends over time (e.g., periods 1-4) in the data, which the model captures in periods 2-4. Since the model is forecasting out-of-sample in sequence blocks 7-9, its ability to forecast those trends in those sequences is particularly noteworthy. Notice that the no-loan rate pops up quite a bit in periods 3-5 in those later

[31] To simplify reporting, we report only the key parameter estimates $(\phi, \delta, \rho, \tau, \alpha, \epsilon)$. The remaining parameter estimates are available from the authors upon request.

[32] Two of the three sessions with low $\hat{\epsilon}$ are sessions with a zero-prior on the honest type, where reputation-building is less common.

sequences, and the model predicts an increase as well. There is less change in the default rate across sequences for the model to capture. It does not pick up the drop in default rate in early periods across sequences well, but it does predict the rate of increase in default rate across periods reasonably well, except for underpredicting reneging in the last period.

4.3. Distinguishing strategic teaching from type-based reputation formation

Strategic teaching generates behavior which is similar to reputation-building in repeated games where there is Bayesian updating of players' unobserved "types". In the type-based models the presence of honest types who prefer to repay creates an incentive (depending on payoffs and the prior P(Honest)) for types with normal payoffs to repay loans.

Many of the predictions of the teaching and type-based models across periods and changes in payoffs are similar. The crucial difference is that in the type-based models a particular *player* has a reputation (i.e., a posterior P(Honest) and associated probabilities of repayment in each period). In the teaching model a *strategy* has a "reputation" or attraction. More precisely, there are four important differences between the teaching and types approaches: Sensitivity to priors, independence of own payoffs and own mixture probabilities, the effect of missed opportunity, and "no second chances". We sketch these differences here and plan to explore them further in future work.

Type-based models have the following properties: If the prior P(Honest) is below some threshold (depending on payoffs and the horizon of the finitely-repeated game) there is no reputation-building; mixture probabilities depend only on the other players' payoffs; if a borrower does not receive a loan in an early period that missed opportunity does not affect future behavior, but if a borrower does not receive a loan in a later period (and hence has no chance to repay and build reputation) then she never receives any more loans; and the sensible sequential equilibrium requires the assumption that if a borrower defaults in one period, then repays in a subsequent period, her reputation is not restored (there are "no second chances"). The teaching model does not make these same predictions. When the type-based models are extended to include quantal-response and a "homemade prior", as in the AQRE model we used as a static benchmark, the sensitivity to priors and independence properties no longer hold, but the missed-opportunity and no-second-chances properties still hold (in a probabilistic sense).

There is a simple experimental way to discriminate between the teaching and type-based approaches. The type-based approach requires that a player's type remain fixed throughout a sequence. If types are drawn randomly in each period, the link between past behavior, inferences about

types, and future incentives is broken and there is no equilibrium reputation-building by normal types. In the teaching approach the presence of nice types *does* matter[33] but it makes little difference whether a player's type is fixed within a sequence or drawn independently in each period. Comparing behavior in experiments with fixed types and independent types therefore provides a way to distinguish the type-based and teaching approaches. The type-based approach predicts a big difference in behavior across those two protocols, while the teaching approach predicts little difference.

We do not suggest that the teaching approach should completely replace type- based equilibrium models of reputation-formation. However, it has always seemed dubious that players are capable of the delicate balance of reasoning required to implement the type-based models, unless they learn the equilibrium through some adaptive process. The teaching model is the only available model of that process, and is therefore worth exploring further. Note also that the type-based models assume optimization and foresight by reputation- builders, and Bayesian updating of types by "learners". The teaching model only changes the last feature, replacing Bayesian updating by learners with learning about their strategies. Our adaptive EWA work showed that some Bayesian learning models which are used to compute expected payoffs (weighted fictitious play) can be perfectly operationalized by generalized reinforcement which keeps track of historical payoffs. In a similar way, assuming that entrants update strategy attractions may be a sensible empirical alternative to Bayesian updating of types, and softens the sharp predictions which result from that approach (particularly the missed opportunity and no-second-chance features).

5. CONCLUSION

This paper extends earlier work on adaptive EWA learning to include sophistication and strategic teaching. Before proceeding to summarize our conclusions, it is helpful to think of the properties one would like an empirical model to have. (i) The model should use all the information that subjects have, if the subjects use that information. (Reinforcement and belief models don't have this property; see Salmon [61].) (ii) The parameters of the model should have psychological interpretations, preferably consistent with accepted ideas in neighboring social sciences. (iii) The model should be as simple as possible, in the sense that every parameter should play a distinct role that is predictively useful. (iv) The model should fit well, both in- and out-of-sample, judging by statistical criteria which, preferably, permit model comparison.

[33]Presence of honest types matters because they alter the attractions of loan and no-loan strategies for lenders, and hence alter the marginal incentives to teach.

EWA does well on all four of these criteria. An important fifth property is that a model be tractable enough to explore its theoretical implications. Heller and Sarin [30] have made initial progress, using a variant of EWA, which is promising. In current work, we have endogenized EWA parameters, making them functions of experience (Ho et al. [31]). Our method opens the door to easier theorizing.[34]

However, adaptive EWA is incomplete by the information-use (i) and psychological fidelity criteria (ii), because it does explain how players' information about the payoffs of others is used, and it does not allow the sort of anticipatory learning which is plausible for intelligent experienced players. Therefore, we extended the model by assuming some fraction α of players are sophisticated in a specific sense: They believe others adapt according to EWA, but also believe that a fraction α' are sophisticated like themselves.

We estimate the sophisticated EWA model on a sample of data from dominance- solvable 'p-beauty contest' games. In these games, each of the n players choose a number from the interval [0,100] and the player whose number is closest to p times the average number wins a fixed prize. We chose these games because there is substantial learning evident in the data, but the adaptive EWA model (and the special cases of reinforcement and weighted fictitious play) fit miserably (see Camerer and Ho [10]).[35]

Introducing sophistication improves fit substantially. More interestingly, we find that the estimated fraction of sophisticated players, α, rises substantially between sessions with inexperienced subjects and those with experienced subjects (who play a second p-beauty contest with a different value of p). This shows that what experience creates is not just learning about the success of strategies, but also *learning about learning*– or increasing sophistication. Players seem to learn that others are adaptive, and learn to "jump ahead" by anticipating changes by others.[36]

Once sophistication is introduced, whether players will be matched together repeatedly or not could matter. Sophisticated players who understand that others are learning will have an incentive to take actions in period t, which "teach" adaptive players how strategies perform, so the sophisticated can earn a higher payoff in period $t+1$ and beyond. (If play-

[34]Specifically, in the one-parameter "EWA Lite" of Ho et al. [31], when an opponent's behavior stabilize the parameters ϕ and δ both converge toward one, which means the learning rule converges toward fictitious play (if $\kappa = 0$). When another player's behavior stabilizes we can therefore apply convergence theorems which are used to show convergence of fictitious play in some settings (see Fudenberg and Levine [25]).

[35]Reinforcement fits particularly badly because $N-1$ players receive no reinforcement, if they do not win, and yet they appear to learn. This poor fit highlights a well-known limit of reinforcement learning– in many-person games where many players earn no payoff, reinforcement learns far too slowly (see Roth and Erev [58] on market games).

[36]A similar point is made, with quite a different model, by Stahl [66]).

ers are rematched in each period, then this kind of teaching motivation disappears.)

Strategic teaching captures the incentive players have, in repeated games with fixed partners, to implement repeated-game strategies which can lead to results that are not stage-game equilibria. We explore this possibility in borrower-lender trust games. In the trust game, borrowers have an incentive to repay loans in early periods, in order to obtain further loans, but toward the end of the eight-period horizon they should quit repaying. We show that a model with adaptive lenders, and sophisticated borrowers who "strategically teach", can explain the basic patterns in these data reasonably well (and consistently better than agent-based quantal response equilibrium).

The teaching approach shows promise for capturing much of the intuition and empirical regularity of reputation- building in repeated games, without using the type-based equilibrium approach. The device of assuming updated types is useful for explaining why lenders are afraid to loan early, and willing to loan late. Sophisticated EWA with strategic teaching produces the same effect more directly– borrowers have an incentive to repay early because they know that lenders will be convinced, not because they believe the borrower's reputations per se, but simply because they learn that loaning in early periods is good.

The teaching model helps resolve a mystery in the experimental literature on repeated games. Basic patterns in the data do go in the direction predicted by sequential equilibrium– viz., borrowers repay more often in early periods of a sequence, and lenders seem to anticipate or learn this, loan more frequently in earlier periods. But changes in treatment variables do not always create predicted changes in behavior (see Neral and Ochs [51] and Jung, Kagel and Levin [37]), subtle predictions of the equilibrium theory are not confirmed, and equilibrium calculations are so complex that it is hard to believe subjects are calculating rather than learning. As a result, Camerer and Weigelt [14] concluded their paper as follows:

> ...the long period of disequilibrium behavior early in these experiments raises the important question of how people learn to play complicated games. The data could be fit to statistical learning models, though new experiments or new models might be needed to explain learning adequately. (pp 27-28).

Strategic teaching is one possible answer to the question they raised almost 15 years ago.

APPENDIX: THE AGENT-QRE MODEL

This appendix describes how we estimate an agent-QRE model on the trust data. Recall that the lender's payoff structure is: No Loan = 10; Loan, Repay = 40, and Loan, Default = -100. The normal borrower's payoff structure is: No Loan = 10, Loan, Repay = 60, Loan, Default = 150. The honest borrower's payoff structure is: No Loan = 10, Loan, Repay = 60, and Loan, Default = 0. We index the game periods by t. We denote the behavioral strategies as follows: $P_L(t)$ is the probability that lender will loan, $P_D(t)$ is the probability that the normal borrower will repay, and $P_H(t)$ is the probability that the honest borrower will repay in period t. Denote the lender's belief that the borrower is honest at time t by q_t. There are 4 parameters to be estimated: the response sensitivity parameters $\alpha_L, \alpha_D, \alpha_H$ as well as the initial belief q_1.

The expected payoff for the lender in period t for no loan is $U_L(NoLoan, t)$ = 10 and for loan is $U_L(Loan, t) = (1-q_t)\cdot[P_D(t)\cdot 40 + [1-P_D(t)]\cdot -100] + q_t \cdot [P_H(t)\cdot 40 + [1-P_H(t)]\cdot -100]$. The expected payoff for the normal borrower for choosing repay in period t is $U_D(repay, t) = 60 + \sum_{O_t}\sum_{r>t}^{8}\{P_L(r|O_t)\cdot [P_D(r|O_t)\cdot 60 + (1-P_D(r|O_t))\cdot 150] + (1-P_L(r|O_t))\cdot 10\}$ and choosing default is $U_D(default, t) = 150 + \sum_{O_t}\sum_{r>t}^{8}\{P_L(r|O_t)[P_D(r|O_t)\cdot 60 + (1-P_D(r|O_t))\cdot 150] + (1-P_L(r|O_t))\cdot 10\}$. The variable O_t is the set of possible future paths of outcomes from time $t+1$ up to time 8 (e.g. a possible path for O_t might be $Repay, Default, Default$ for $t = 5$.). $P_L(r|O_t)$ is the conditional probability of Loan at time r given the path O_t and there are conditional beliefs $q_{r|O_t}$ associated with these conditional probabilities. We assume that the borrower would not consider any future path in which he will go back to repay after turning default (e.g. $default, repay, default$ would not be in the set of O_5.). The expected payoff for the honest borrower can be defined the same way.

Next, we propose the following procedure to update the belief in time t. If we observe a repay in period $t-1$, we update q_t as follows:

$$q_t = \frac{q_{t-1}\cdot P_H(t-1)}{q_{t-1}\cdot P_H(t-1) + (1-q_{t-1})\cdot P_D(t-1)} \tag{A.1}$$

Otherwise, we update q_t as follows:

$$q_t = \frac{q_{t-1}\cdot (1-P_H(t-1))}{q_{t-1}\cdot (1-P_H(t-1)) + (1-q_{t-1})\cdot (1-P_D(t-1))} \tag{A.2}$$

We also update the conditional beliefs $q_{r|O_t}$ with conditional probabilities using the same procedure above. Note that we need to solve for a set of conditional beliefs $q_{r|O_t}$ that are consistent with all future paths we consider. For example, if two future paths share the same outcomes up to

time r', then both paths should share the same set of conditional belief $q_{r'|O_t}$ up to r'. Given that the conditional beliefs $q_{r|O_t}$ are nonlinear in $P_D(r|O_t), P_H(r|O_t), P_L(r|O_t)$, we find the consistent set of $q_{r|O_t}$ numerically.

REFERENCES

1. W. Amaldoss, "Collaborating to Compete: A Game-theoretical Model and Experimental Investigation of Competition among Alliances," Unpublished Ph.D. dissertation, University of Pennsylvania, 1998.
2. C. Anderson and C. F. Camerer,"Experience-weighted Attraction Learning in Sender-receiver Signaling Games," *Economic Theory*, 16(3), (2001), 689-718.
3. J. Andreoni and J. Miller, "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence," *Economic Journal*, 103 (1993), 570-585.
4. A. Blume, D. DeJong, G. Neumann and N. Savin, "Learning in Sender-receiver Games," University of Iowa Working Paper, 1999.
5. G. W. Brown, "Iterative Solution of Games by Fictitious Play," in *Activity Analysis of Production and Allocation*, John Wiley & Sons, New York, 1951.
6. G. P. Cachon and C. F. Camerer, "Loss-avoidance and Forward Induction in Experimental Coordination Games," *The Quarterly Journal of Economics*, 111 (1996), 165-194.
7. C. F. Camerer, "Progress in Behavioral Game Theory," *Journal of Economic Perspectives*, 11 (1997), 167-188.
8. C.F. Camerer and T-H Ho, "EWA Learning in Normal-form Games: Probability Rules, Heterogeneity and Time Variation," *Journal of Mathematical Psychology*, 42 (1998), 305-326.
9. C.F. Camerer and T-H Ho, "Experience-weighted Attraction Learning in Games: Estimates from Weak-link Games," in *Games and Human Behavior: Essays in honor of Amnon Rapoport*, ed. by D. Budescu, I. Erev, and R. Zwick., Lawrence Erlbaum Assoc. Inc., New Jersey, 1999, 31-52.
10. C. F. Camerer and T-H Ho, "Experience-weighted Attraction Learning in Normal-form Games," *Econometrica*, 67 (1999), 827-874.
11. C. F. Camerer, T-H Ho, and X. Wang, "Individual Differences and Payoff Learning in Games," University of Pennsylvania Working Paper, 1999.
12. C. F. Camerer, D. Hsia, and T-H Ho, "EWA Learning in Bilateral Call Markets," University of Pennsylvania Working Paper, 2000.
13. C. F. Camerer, E. Johnson, S. Sen and T. Rymon, "Cognition and Framing in Sequential Bargaining for Gains and Losses," in *Frontiers of Game Theory*, ed. by K Binmore, A. Kirman, and P. Tani., MIT Press, Cambridge, 27-48, 1993.
14. C. F. Camerer and K. Weigelt, "Experimental Tests of A Sequential Equilibrium Reputation Model," *Econometrica*, 56 (1988), 1-36.
15. H. Chen, J. W. Friedman, and J. F. Thisse, "Boundedly Rational Nash Equilibrium: A Probabilistic Choice Approach," *Games and Economic Behavior*, (in press).
16. Y. Chen and Y. Khoroshilov, "Learning Under Limited Information," University of Michigan Working Paper, 2000.
17. K. Clark and M. Sefton, "Matching Protocols in Experimental Games," University of Manchester Working Paper, 1999.

18. D. J. Cooper and C. Stockman, "Fairness, Learning and Constructive Preferences: An Experimental Examination," Case Western Reserve University Working Paper, 1999.

19. M. Costa-Gomes, V. Crawford, and B. Broseta, "Cognition and Behavior in Normal-form Games: An Experimental Study," *Econometrica* (in press).

20. J. Duffy and N. Feltovich, "Does Observation of Others Affect Learning in Strategic Environments? An Experimental Study," *International Journal of Game Theory*, 28 (1999), 131-152.

21. J. Duffy, and R. Nagel, "On the Robustness of Behavior in Experimental 'Beauty Contest' Games," *Economic Journal*, 107 (1997), 1684-1700.

22. G. Ellison, "Learning from Personal Experience: One Rational Guy and the Justification of Myopia," *Games and Economic Behavior*, 19 (1997), 180-210.

23. D. Engelmann and M. Strobel, "The False Consensus Effect Disappears if Representative Information and Monetary Incentives are Given," *Experimental Economics*, 3 (2001), 241-260.

24. I. Erev and A. Roth, "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed-strategy Equilibria," *The American Economic Review*, 88 (1998), 848-881.

25. D. Fudenberg and D. Levine, "Reputation and Equilibrium Selection in Games with A Patient Player," *Econometrica*, 57 (1989), 759-778.

26. D. Fudenberg and D. Levine, *The Theory of Learning in Games*, MIT Press, Cambridge, 1998.

27. J. Gale, K. Binmore, and L. Samuelson, "Learning to be Imperfect: The Ultimatum Game," *Games and Economic Behavior*, 8 (1995), 56-90.

28. R. Gallistel, *The Organization of Learning*, MIT Press, Cambridge, 1990.

29. J.K. Goeree and C. A. Holt, "Stochastic Game Theory: For Playing Games, Not Just for Doing Theory," *Proceedings of the National Academy of Sciences*, 96 (1999), 10564-10567.

30. D. Heller and R. Sarin, "Parametric Adaptive Learning," *Econometric Society World Congress 2000 Contributed Papers*, No. 1496, 2000.

31. T-H Ho, C. F. Camerer, and J. K. Chong, "Economic Value of EWA Lite: A Functional Theory of Learning in Games," University of Pennsylvania Working Paper, 2001.

32. T-H Ho, C. F. Camerer, and K. Weigelt, "Iterated Dominance and Iterated Best-response in p-Beauty Contests," *American Economic Review*, 88 (1998), 947-969.

33. T-H Ho and J. K. Chong, "A Parsimonious Model of SKU Choice," University of Pennsylvania Working Paper, 1999.

34. C. A. Holt, "Industrial Organization: A Survey of Laboratory Research," in *Handbook of Experimental Economics* ed. by J. Kagel and A. Roth, Princeton: Princeton University Press, Princeton, 1995, 349-443.

35. E. Hopkins, "Two Competing Models of How People Learn in Games," University of Edinburgh and Pittsburgh Working Paper, 1999.

36. D. Hsia, "Learning in Single Call Markets," Unpublished Ph.D. Dissertation, University of Southern California, 1999.

37. Y. J. Jung, J. H. Kagel, and D. Levin, "On the Existence of Predatory Pricing: An Experimental Study of Reputation and Entry Deterrence in the Chain-store Game," *RAND Journal of Economics*, 25 (1994), 72-93.

38. M.J. Knez and C. F. Camerer, "Increasing Cooperation in Prisoner's Dilemma by Establishing a Precedent of Cooperation in Coordination Games," *Organizational Behavior and Human Decision Processes*, (in press).

39. E. Kohlberg and J-F Mertens, "On the Strategic Stability of Equilibria,", *Econometrica*, 54 (1986), 1003-1038.

40. D. Kreps and R. Wilson, "Reputation and Imperfect Information," *Journal of Economic Theory*, 27 (1982), 253-279.

41. P. H. McAllister, "Adaptive approaches to stochastic programming," *Annals of Operations Research*, 30 (1991), 45-62.

42. R. D. McKelvey and T. R. Palfrey, "An Experimental Study of the Centipede Game," *Econometrica*, 60 (1992), 803-836.

43. R. D. McKelvey and T. R. Palfrey, "Quantal Response Equilibria for Normal-form Games," *Games and Economic Behavior*, 7 (1995), 6-38.

44. R. D. McKelvey and T. R. Palfrey, "Quantal Response Equilibria for Extensive-form Games," *Experimental Economics*, 1 (1998), 9-41.

45. R. D. McKelvey and G. Riddough, "Is Economics the Hard Science?" Caltech Working Paper, 1999.

46. D. Mookerjee and B. Sopher, "Learning Behavior in an Experimental Matching Pennies Game," *Games and Economic Behavior*, 7 (1994), 62-91.

47. D. Mookerjee and B. Sopher, "Learning and Decision Costs in Experimental Constant-sum Games," *Games and Economic Behavior*, 19 (1997), 97-132.

48. J. Morgan and M. Sefton, "An Empirical Investigation of Unprofitable Games," Princeton University Working Paper, 1999.

49. R. Nagel, "Experimental Results on Interactive Competitive Guessing," *The American Economic Review*, 85 (1995), 1313-1326.

50. R. Nagel, "A Review of Beauty Contest Games," in *Games and Human Behavior: Essays in honor of Amnon Rapoport*, ed. by D. Budescu, I. Erev, and R. Zwick., Lawrence Erlbaum Assoc. Inc., New Jersey, 1999, 105-142.

51. J. Neral and J. Ochs, "The Sequential Equilibrium Theory of Reputation Building: A Further Test," *Econometrica*, 60 (1992), 1151-1169.

52. J. Ochs, "Entry in Experimental Market Games," in *Games and Human Behavior: Essays in honor of Amnon Rapoport*, ed. by D. Budescu, I. Erev, and R. Zwick., Lawrence Erlbaum Assoc. Inc., New Jersey, 1999.

53. T. R. Palfrey and H. Rosenthal, "Private Incentives in Social Dilemmas: The Effects of Incomplete Information and Altruism," *Journal of Public Economics*, 35 (1988), 309-332.

54. J. Partow and A. Schotter, "Does Game Theory Predict Well for the Wrong Reasons? An Experimental Investigation," C.V. Starr Center for Applied Economics Working Paper 93-46, New York University, 1993.

55. A. Rapoport and W. Amaldoss, "Mixed Strategies and Iterative Elimination of Strongly Dominated Strategies: An Experimental Investigation of States of Knowledge," *Journal of Economic Behavior and Organization*, 42 (2000), 483-521.

56. A. Rapoport, A. K-C Lo, and R. Zwick, "Choice of Prizes Allocated by Multiple Lotteries with Endogenously Determined Probabilities," University of Arizona, Department of Management and Policy working paper, 1999.

57. R. W. Rosenthal, "Games of Perfect Information, Predatory Pricing and Chain-Store Paradox," *Journal of Economic Theory*, 25 (1981), 92-100.

58. A. Roth and I. Erev, "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, 8 (1995), 164-212.

59. R. Sarin and F. Vahid, "Strategy Similarity and Coordination," Texas A&M University Working Paper, 2000.

60. R. Sarin and F. Vahid, "Predicting How People Play Games: A Simple Dynamic Model of Choice," *Games and Economic Behavior*, 34 (2001), 104-122.

61. T. Salmon, "Evidence for 'Learning to Learn' Behavior in Normal-form Games," Caltech Working Paper, 1999.

62. T. Salmon, "An Evaluation of Econometric Models of Adaptive Learning," *Econometrica*, (in press).

63. R. Selten, "Anticipatory learning in 2-person Games," University of Bonn Discussion Paper Series B, 1986.

64. R. Selten, "Evolution, Learning, and Economic Behavior," *Games and Economic Behavior*, 3 (1991), 3-24.

65. R. Selten and R. Stoecker, "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames: A Learning Theory Approach," *Journal of Economic Behavior and Organization*, 7 (1986), 47-70.

66. D. O. Stahl, "Boundedly Rational Rule Learning in a Guessing Game," *Games and Economic Behavior*, 16 (1996), 303-330.

67. D. O. Stahl, "Rule Learning in Symmetric Normal-form Games: Theory and Evidence," *Games and Economic Behavior*, 32 (2000), 105-138.

68. D. O. Stahl, "Population Rule Learning in Symmetric Normal-form Games: Theory and Evidence," *Journal of Economic Behavior and Organization*, 45 (2001), 19-35.

69. D. O. Stahl, and P. Wilson, "On Players Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10 (1995), 213-54.

70. A. Tversky and D. Kahneman, "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *The Quarterly Journal of Economics*, 106 (1991), 1039-1061.

71. J. Van Huyck, R. Battalio, and R. Beil, "Tacit Cooperation Games, Strategic Uncertainty, and Coordination Failure," *The American Economic Review*, 80 (1990), 234-248.

72. J. Watson, "A 'Reputation' Refinement without Equilibrium," *Econometrica*, 61 (1993), 199-205.

73. J. Watson and P. Battigali, "On 'Reputation' Refinements with Heterogeneous Beliefs," *Econometrica*, 65 (1997), 363-374.

74. G. Weizsacker, "Ignoring the Rationality of Others: Evidence from Experimental Normal-form Games," Harvard Business School Working Paper, 2000.

Figure 1: The EWA Learning Cube

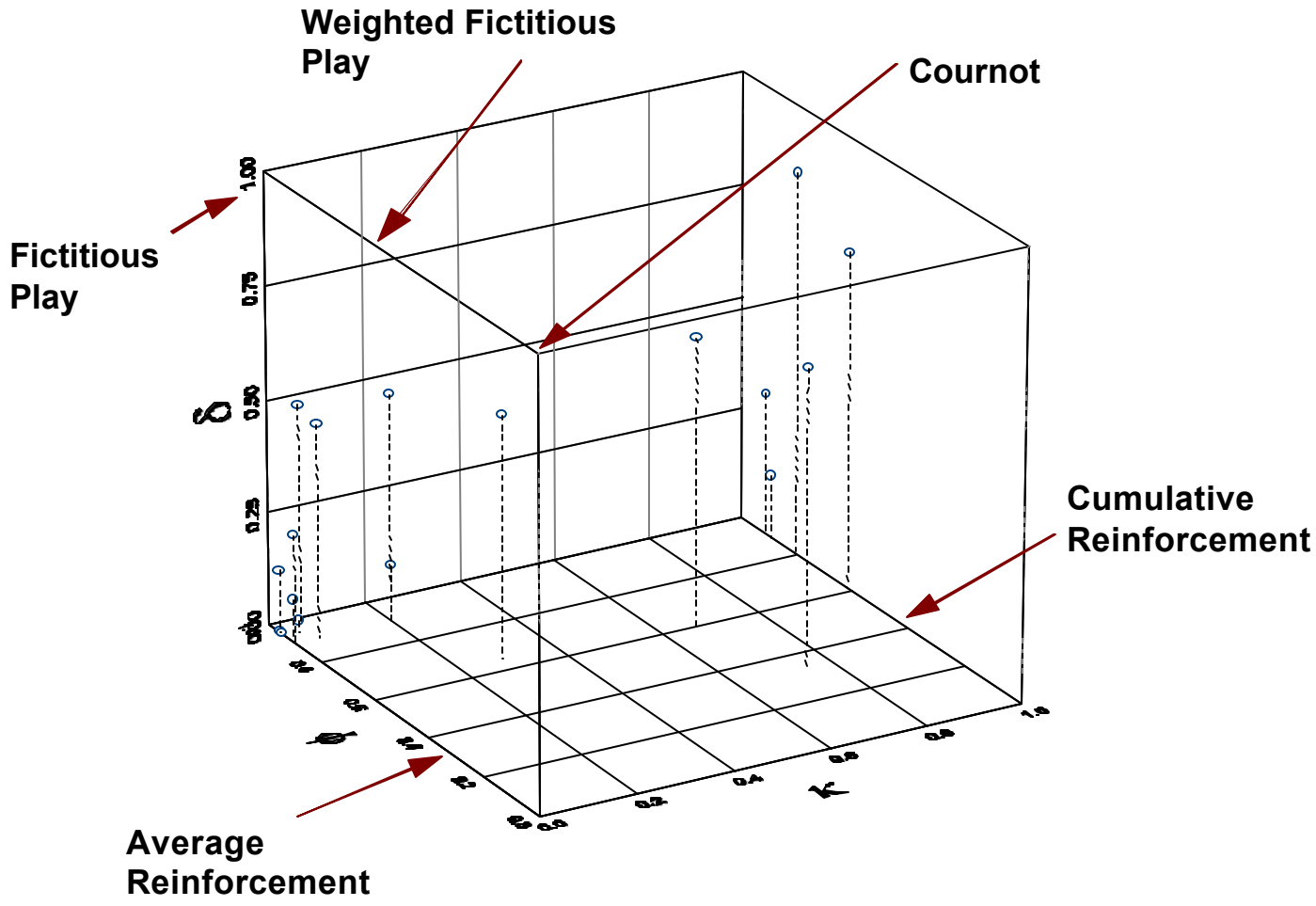


Figure 2a: Actual Choice Frequencies for Inexperienced Subjects

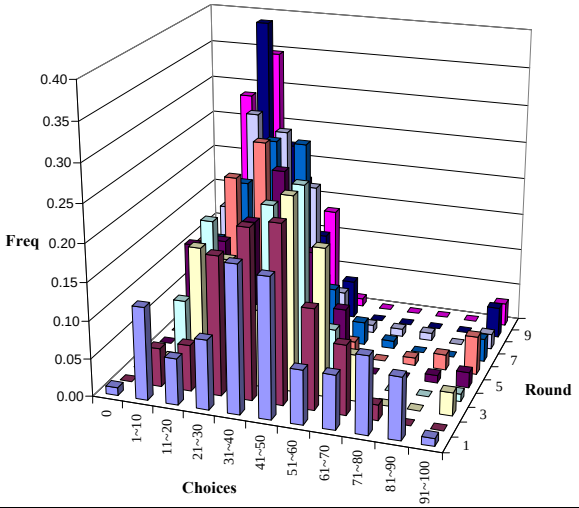


Figure 2b: Adaptive EWA Model Frequencies for Inexperienced Subjects

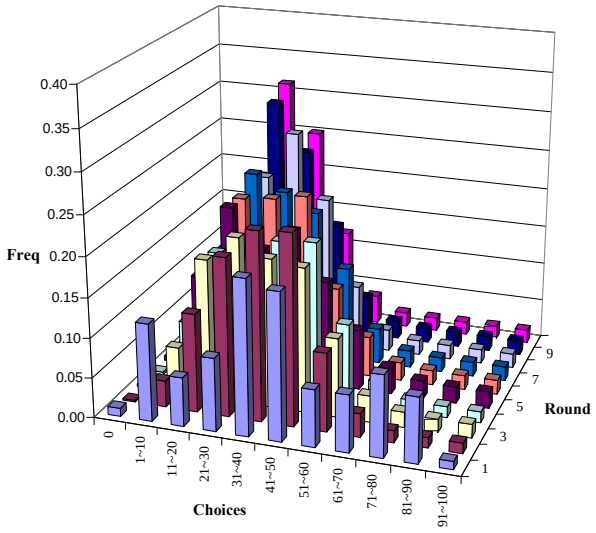


Figure 2c: Sophisticated EWA Model Frequencies for Inexperienced Subjects

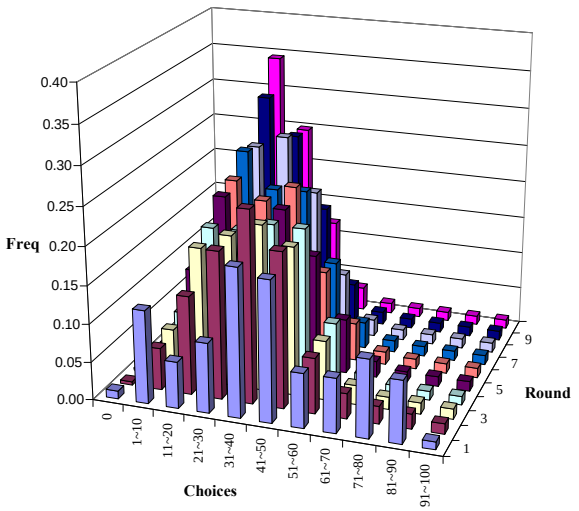


Figure 3a: Actual Choice Frequencies for Experienced Subjects

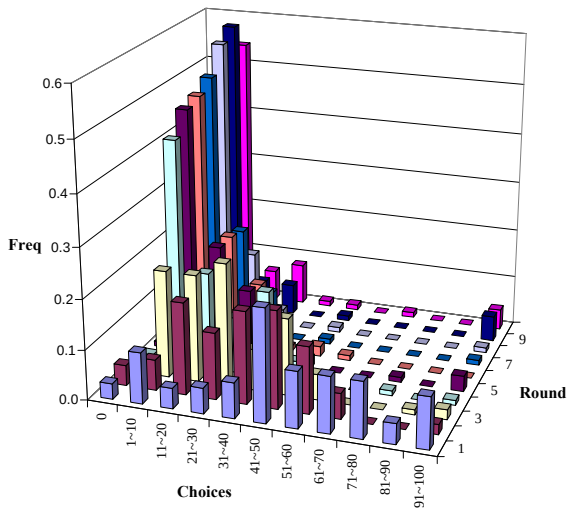


Figure 3b: Adaptive EWA Model Frequencies for Experienced Subjects

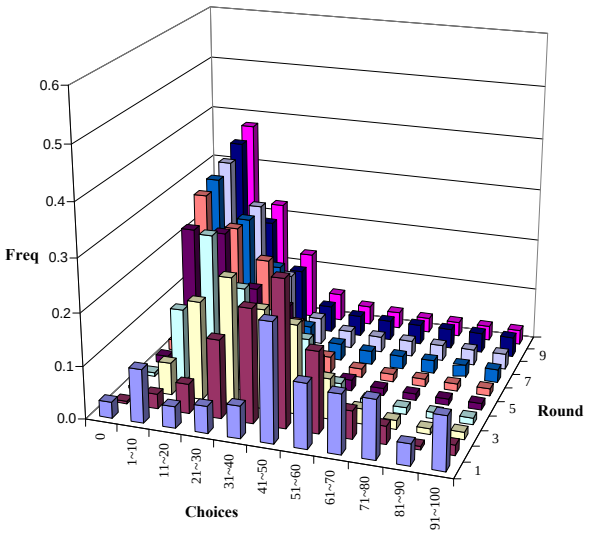


Figure 3c: Sophisticated EWA Model Frequencies for Experienced Subjects

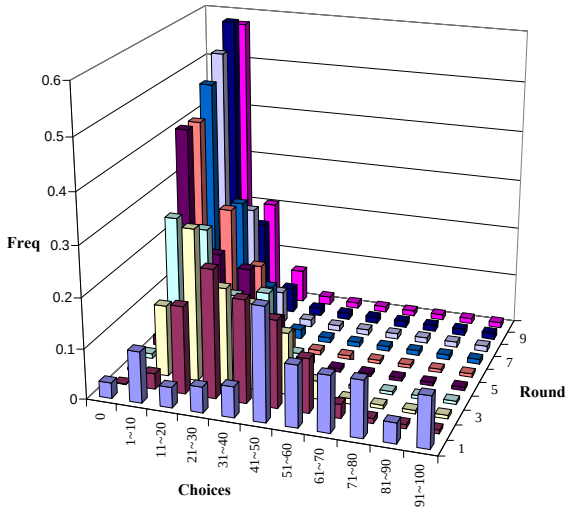


Figure 4: Borrower's optimal mixed-strategy of pay back as predicted by sequential equilibrium

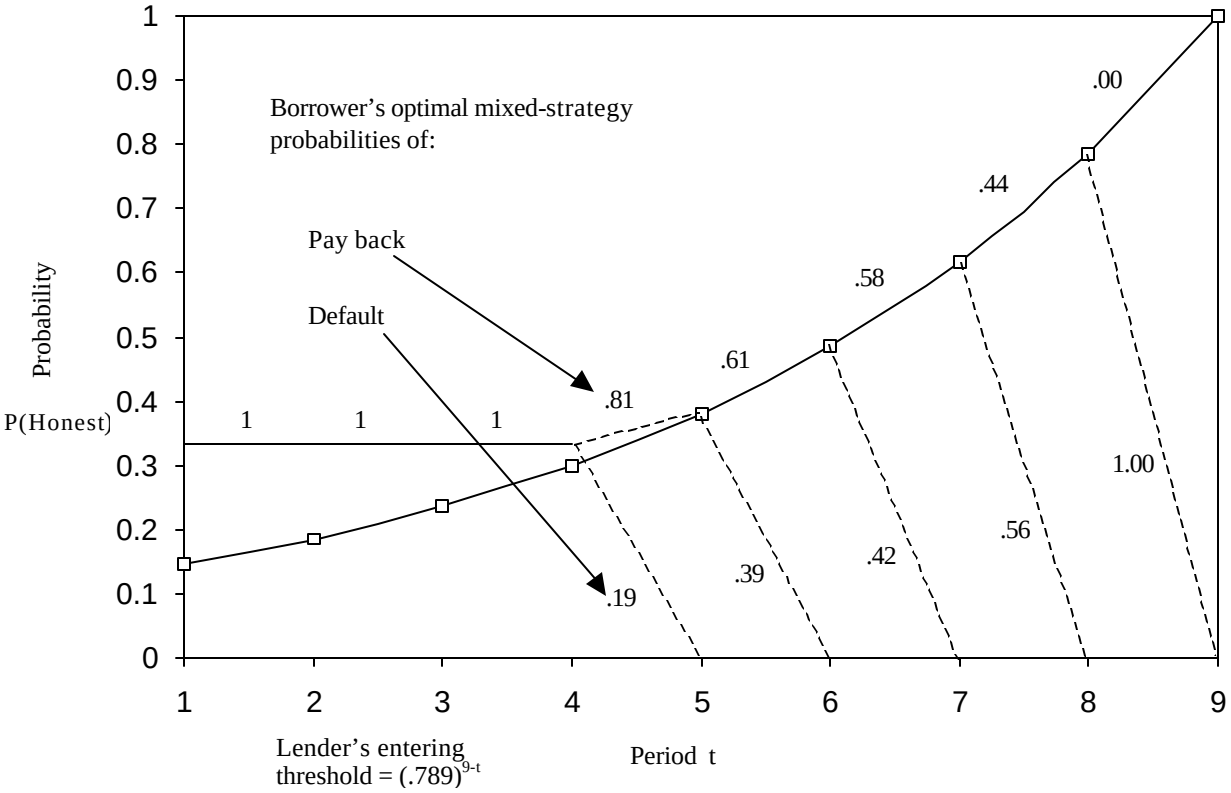


Figure 5a: Empirical Frequency for No Loan

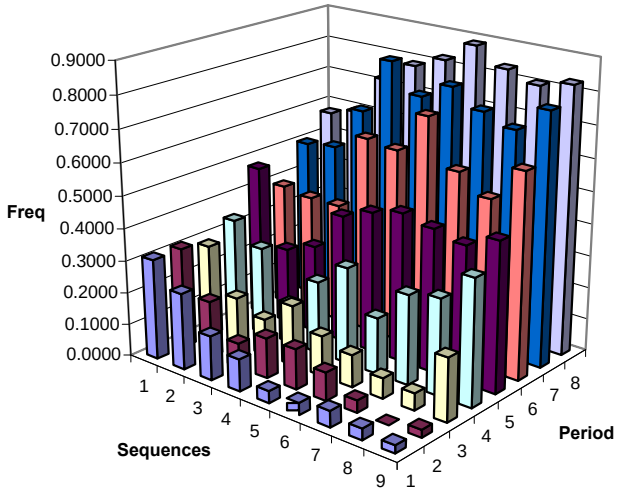


Figure 5c: Predicted Frequency for No Loan

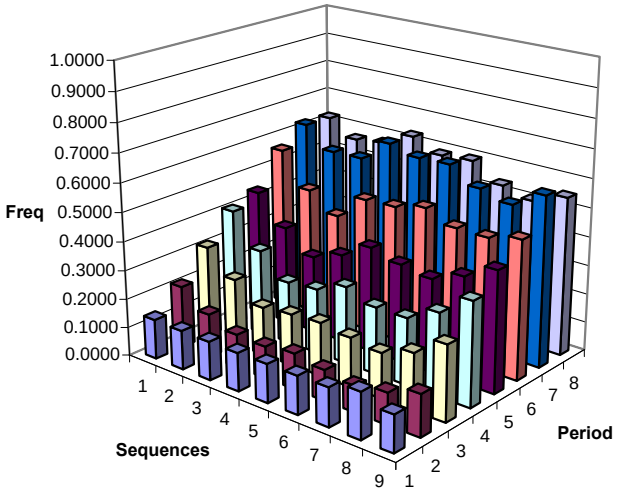


Figure 5b: Empirical Frequency for Default conditional on Loan

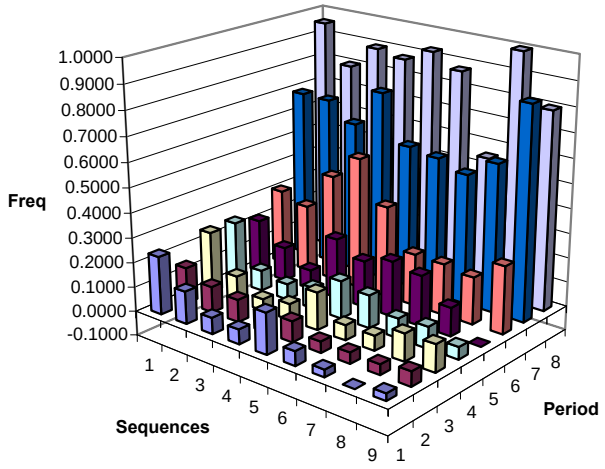


Figure 5d: Predicted Frequency for Default conditional on Loan

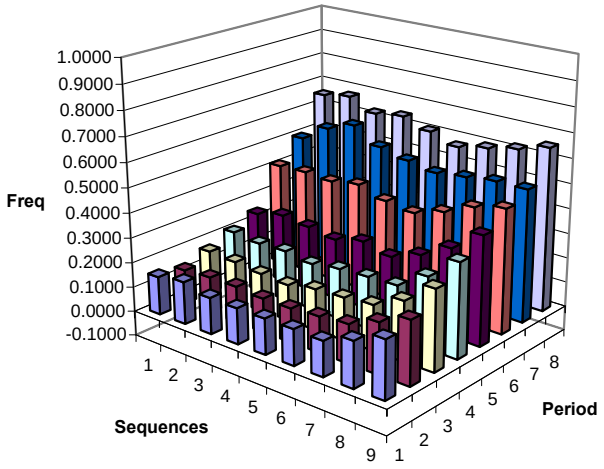


Table 1a: A summary of EWA parameter estimates and forecast accuracy (games estimated by us)

		EWA estimates (standard error)			Model accuracy					Comments
CITATION	GAME	δ	ϕ	$\rho=(1-\kappa)\phi$	EWA	Choice reinforce – EWA	Belief – EWA	In / Out of sample	Fit tech-nique	
Camerer, Ho and Hsia (2000)	Sealed bid mechanism+	n.a.	1.00	0.91	1102.0	30.8	65.5	IN	-LL	ψ, ω & κ replace δ & ρ
Camerer, Ho and Wang (1999)	"Continental divide" coordination	0.75	0.61	0.00	346.9	86.1	235.8	OUT	-LL	
Camerer and Ho (1998)	Weak-link coordination	0.65	0.58	0.20	358.1	29.1	438.6	IN	-LL	
Anderson and Camerer (in press)	Signaling games (game 3) 95% Confidence Interval	0.69 (0.47,1.00)	1.020 (0.99,1.04)	1.00 (0.98,1.00)	72.2	6.5	10.1	OUT	-LL	
	Signaling games (game 5) 95% Confidence Internal	0.54 (0.45,0.63)	0.65 (0.59,0.71)	0.46 (0.39, 0.54)	139.5	14.1	23.7	OUT	-LL	
Camerer and Ho (1999b)	Median-action coordination	0.85 (0.01)	0.80 (0.02)	0.00 (0.00)	41.1	39.2	72.8	OUT	-LL	
	4x4 Mixed-strategy games	0.00 (0.04)	1.04 (0.01)	0.96 (0.01)	326.4	9.1	-40.8	OUT	-LL	Payoff = 5 rupees
		0.73 (0.10)	1.01 (0.01)	0.95 (0.01)	341.7	18.0	8.4	OUT	-LL	Payoff =10 rupees
	6x6 Mixed-strategy games	0.41 (0.08)	0.99 (0.01)	0.94 (0.01	301.7	6.8	-5.4	OUT	-LL	Payoff =5 rupees
		0.55 (0.05)	0.99 (0.01)	0.93 (0.02)	362.3	13.7	8.9	OUT	-LL	Payoff =10 rupees
	p-beauty contests*	0.95 (0.01)	0.11 (0.00)	0.00 (0.00)	1917.0	647.0	35.0	OUT	-LL	Experienced and Inexperienced Combined
Camerer, Ho and Wang (1999)	Normal form centipede (odd player)	0.32 (0.32)	0.91 (0.14)	0.00 (0.00)	1016.8	57.6	536.3	OUT	-LL	Clairvoyance full update, κ
	Normal form centipede (even player)	0.24 (0.32)	0.90 (0.14)	0.95 (0.03)	951.3	46.4	604.7	OUT	-LL	Clairvoyance full update, κ

+ In Figure 1, we did not include this study.

* Unlike the previous estimates, these new estimates assume that subjects do not know the winning numbers.

Table 1b: A summary of EWA parameter estimates and forecast accuracy (games estimated by others)

		EWA estimates (standard error)			Model accuracy					Comments
CITATION	GAME	δ	ϕ	$\rho=(1-\kappa)\phi$	EWA	Choice reinforce – EWA	Belief – EWA	In / Out of sample	Fit tech-nique	
Chen and Khoroshilov (1999)	Cost allocation+	.80~1.0	1 (fixed)	.1~.3	.73~.88	-.01~.07	n.a.	IN	MSD	
Morgan and Sefton (1999)	"Unprofitable" games (baseline games)	0.08 (0.07)	0.93 (0.01)	0.92 (0.03)	1729.5	0.6	n.a.	IN	-LL	
	"Unprofitable" games (upside games)	0.14 (0.06)	0.89 (0.01)	0.00 (0.00)	1906.5	16.2	n.a.	IN	-LL	
Stahl (1999)	5x5 matrix games	.66 (0.02)	.34 (0.04)	.086 (0.08)	4803.7	64.7	n.a.	OUT	-LL	
Hsia (1999)	Call markets	0.47 (0.32)	0.97 (0.01)	0.74 (0.06)	1915.0	0.0	403.0	IN	-LL	
Amaldoss (1998)	Same function alliance – equal profit sharing	0.00	0.95	0.89	886.3	1.6	529.6	IN	-LL	High reward
		0.00	0.99	0.99	767.5	30.1	390.4	IN	-LL	Med. Reward
		0.07	0.89	0.87	1399.7	9.4	541.5	IN	-LL	Low reward
	Same function alliance – proportional sharing	0.00	0.99	1.00	910.8	36.4	813.0	IN	-LL	High reward
		0.00	0.92	0.96	1055.0	18.3	615.9	IN	-LL	Med. Reward
		0.00	0.97	0.96	1013.7	13.3	1095.6	IN	-LL	Low reward
	Parallel development of product – equal sharing	0.00	0.91	0.59	1194.2	0.1	566.3	IN	-LL	High reward
		0.17	0.90	0.93	1321.5	9.5	497.2	IN	-LL	Med. Reward
		0.21	0.88	0.66	1297.7	4.5	484.1	IN	-LL	Low reward
Rapoport and Amaldoss (2000)	Patent race game – symmetric players	0.00	0.94	0.93	3551.7	12.1	1097.7	IN	-LL	Low reward
		0.00	0.97	0.98	2908.1	20.2	725.9	IN	-LL	High reward
	Patent race game – asymmetric players	0.48	0.90	0.86	3031.5	89.1	706.8	IN	-LL	Strong player
		0.14	0.96	0.97	2835.5	15.7	611.0	IN	-LL	Weak player

+ In Figure 1, we did not include this study.

Table 2. Model Parameter Estimates for p-beauty Contest Game

	INEXPERIENCED SUBJECTS			EXPERIENCED SUBJECTS		
	Sophisticated EWA	Adaptive EWA	QRE[1]	Sophisticated EWA	Adaptive EWA	QRE
ϕ	**0.44**	0.00	-	**0.29**	0.22	-
	(0.05)[2]	(0.00)	-	(0.03)	(0.03)	-
δ	**0.78**	0.90	-	**0.67**	0.99	-
	(0.08)	(0.05)	-	(0.05)	(0.02)	-
ρ	**0.00**	0.00	-	**0.01**	0.00	-
	(0.00)	(0.00)	-	(0.00)	(0.00)	-
α	**0.24**	0.00	1.00	**0.77**	0.00	1.00
	(0.04)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)
α'	**0.00**	0.00	-	**0.41**	0.00	-
	(0.00)	(0.00)	-	(0.03)	(0.00)	-
d	**0.16**	0.13	0.04	**0.15**	0.11	0.04
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
LL						
(in sample)	**-2095.32**	-2155.09	-2471.50	**-1908.48**	-2128.88	- 2141.45
(out of sample)	**-968.24**	-992.47	-1129.25	**-710.28**	-925.09	- 851.31
Avg. Prob.						
(in sample)	6%	5%	3%	7%	5%	5%
(out of sample)	7%	7%	4%	13%	9%	9%

[1] We also estimated the QRE model with different λs in each period. For inexperienced players, the λs in period 2 to 6 are: 0.590; 0.663; 0.941; 1.220; 1.221; 1.381. In sample LL=-2460.51; out of sample LL=-1100.09. For experienced players, the λ's are 1.382; 2.627; 3.970; 5.249; 5.363; 8.399. In sample LL=-2074.25; out of sample LL=-769.19.

[2] Standard errors in parenthesis.

Table 3: Payoffs in the borrower-lender game, Camerer & Weigelt (1988)

Lender strategy	Borrower strategy	payoffs to lender	payoffs to borrower	
			normal (X)	honest (Y)
loan	default	-100^{*}	150	0
	repay	40	60	60
no loan	no choice	10	10	10

Note: * Loan-default lender payoffs were -50 in sessions 6-8 and -75 in sessions 9-10.

Table 4: A Comparison of In-Sample and Out-of-Sample Performance Between the Teaching and AQRE Models

Experiment No:	3	4	5	6	7	8	9	10
No of Sequence (Total):	90	90	81	70	77	69	90	101
In-Sample Calibration								
The Teaching Model								
Average Probability	75%	74%	75%	83%	75%	80%	59%	77%
Log-likelihood	-268.85	-312.64	-277.98	-178.58	-281.96	-201.84	-443.21	-331.11
Agent-based Quantal Response Equilibrium (AQRE)								
Average Probability	67%	67%	71%	82%	67%	73%	54%	76%
Log-likelihood	-325.48	-350.97	-297.28	-190.02	-322.03	-247.49	-454.68	-320.52
Out-Of-Sample Validation								
The Teaching Model								
Average Probability	73%	73%	76%	86%	72%	78%	59%	79%
Log-likelihood	-139.29	-141.04	-131.38	-69.60	-137.72	-98.46	-249.41	-137.50
Agent-based Quantal Response Equilibrium (AQRE)								
Average Probability	70%	69%	74%	85%	66%	73%	54%	73%
Log-likelihood	-144.58	-149.20	-126.32	-72.24	-159.49	-115.55	-270.72	-170.53

Note: Data source from Camerer and Weigelt (1988)

Table 5: Parameter Estimates of the Teaching Model

Experiment No.:	3	4	5	6	7	8	9	10
No of Sequence (Total):	90	90	81	70	77	69	90	101
Parameters for Adaptive Lender								
ϕ	0.29	0.64	0.71	0.72	0.55	0.68	0.69	0.09
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.04)	(0.07)	(0.02)
δ	0.34	0.19	0.34	0.88	0.18	0.65	0.40	0.49
	(0.04)	(0.06)	(0.17)	(0.26)	(0.08)	(0.14)	(0.10)	(0.06)
κ	0.99	0.88	0.36	0.40	1.00	0.12	0.39	0.88
	(0.02)	(0.16)	(0.06)	(0.08)	(0.06)	(0.08)	(0.15)	(0.84)
τ	0.51	0.97	0.96	0.88	0.94	0.97	1.00	0.96
	(0.02)	(0.06)	(0.04)	(0.11)	(0.06)	(0.04)	(0.05)	(0.06)
Parameters for Adaptive Borrower								
ϕ	1.00	0.69	0.72	0.52	0.93	0.69	0.10	0.72
	(0.15)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.11)	(0.03)
δ	0.17	0.44	0.79	0.48	0.47	0.55	0.16	0.33
	(0.17)	(0.03)	(0.06)	(0.05)	(0.02)	(0.03)	(0.18)	(0.06)
κ	0.90	0.89	0.67	0.73	0.58	0.83	0.82	0.58
	(0.21)	(0.06)	(0.21)	(0.20)	(0.11)	(0.14)	(0.17)	(0.12)
τ	0.10	0.36	0.55	0.11	0.55	0.35	0.25	0.43
	(0.23)	(0.08)	(0.07)	(0.05)	(0.09)	(0.05)	(0.31)	(0.09)
Parameters for Sophisticated Borrower								
ε	1.00	1.00	0.95	0.93	0.22	0.93	0.19	0.25
	(0.01)	(0.28)	(0.09)	(0.21)	(0.21)	(0.27)	(0.13)	(0.17)
α	0.70	0.26	0.14	0.03	0.07	0.28	0.99	0.03
	(0.08)	(0.08)	(0.07)	(0.03)	(0.06)	(0.08)	(0.30)	(0.03)

Note: Each parameter is presented with its standard error (in parenthesis) directly below.